L113 Word Meaning and Discourse Understanding

Session 8: Discourse Theories

Simone Teufel

Natural Language and Information Processing (NLIP) Group
UNIVERSITY OF CAMBRIDGE
Simone.Teufel@cl.cam.ac.uk

2011/2012

What is meant by Discourse Structure?

- Definition of discourse structure: "the organisation of language above the sentence"
- We have so far talked about the: local structure of discourse
  - Anaphoric structure
  - Entity structure
  - Topical structure (lexical chains)
  - Temporal structure
- How could discourse be structured more globally?
  - Intentional structure
  - Rhetorical structure
  - Argument structure
- Discourse theories can be text type and domain inspecific (text type = novel, scientific article etc; domain = history vs. psycholinguistics)

Grosz and Sidner (1986)

- "Attention, Intentions, and the Structure of Discourse" (CL, 1986)
  - Linguistic structure: hierarchical discourse segments
  - Intentional structure: **speaker-centric**; communicative purpose of segments and relations between the purposes
  - Attentional state: **listener-centric**; salient objects, properties, relations (and how they help them keeping track of referring expressions in discourse)
- Model explains cue phrases, referring expressions, interruptions
- Can be used to infer discourse structure if there is knowledge about reference or vice versa

Linguistic structure

DS1	1	E	First you have to remove the flywheel.
	2	A	How do I remove the flywheel?
	3	E	First, loosen the two allen head setscrews holding it to the shaft, then pull it off.
	4	A	OK.
DS2	5		I can only find one screw. Where is the other one?
	6	E	On the hub of the flywheel.
	7	A	That is the one I found. Where is the other one?
	8	E	Around 90 degrees around the hub from the first one.
	9	A	I don't understand. I can only find one. Oh wait, yes, I think I was on the wrong wheel.
	10	E	Show me what you are doing.
	11	A	I was on the wrong wheel and I find them both now.
DS3	12		The tool I have is awkward. Is there another tool that I could use instead?
	13	E	Show me the tool you are using.
	14	A	OK.
	15	E	Are you sure you are using the right size key?
	16	A	I'll try some others.
	17		I found an angle I can get at it.
DS4	18		The two screws are loose, but I am having trouble getting the wheel off.
DS5	19	E	Use the wheelpuller. Do you know how to use it?
	20	A	No.
	21	E	Do you know what it looks like?
	22	A	Yes.
	23	E	Show it to me please.
	24	A	OK.
	25	E	Good. Loosen the screw in the center and place the jaws around the hub of the wheel, then tighten the screw onto the center of the shaft. The wheel should slide off.

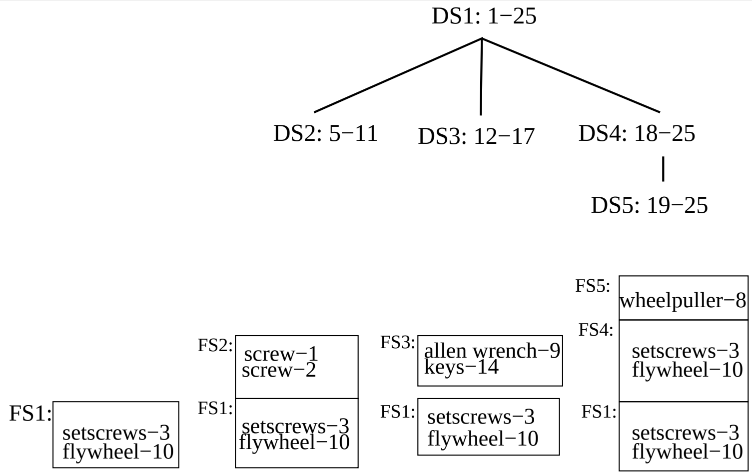
Intentional structure

- Intentions associated with discourse segments:

  I1: (Intend E (Intend A (Remove A flywheel)))
  I2: (Intend A (Intend E (Tell E A (Location other setscrew))))
  I3: (Intend A (Intend E (Identify E A another tool)))
  I4: (Intend A (Intend E (Tell E A (How (Getoff A wheel)))))
  I5: (Intend E (Know-How-To A (Use A wheelpuller)))

- Two structural relations hold between the segments:
  - Dominance: DSP1 dominates DSP2 ⇔ An action that satisfies intention DSP2 is intended to provide part of the satisfaction of intention DSP1
    I1 DOM I2 I1 DOM I4 I1 DOM I3 I4 DOM I5
  - Satisfaction-precedence: DSP1 satisfaction-precedes DSP2 ⇔ DSP1 must be satisfied before DSP2:
    I2 SP I3 I2 SP I4 I3 SP I4

Attentional Structure



- Dynamic attentional state records salient objects, properties and relations for each point in the conversation
- Relationships between intentional segments determine pushes and pops in the focus spaces
- Claim: focus structure constrains use of linguistic expressions

Grosz and Sidner (1986): problems

- Theory strongly influenced by analysis of spoken language
- Text type of expert-apprentice conversations:
  - Underlying hierarchical task-structure, unlike in general conversations/texts
  - This task-structure provides common knowledge about the task
- Theory requires
  - Recognition of intentions in text (AI-complete problem)
  - Representation of participants' knowledge of the domain

$\rightarrow$ computational feasibility?

Rhetorical discourse processing and information access

- For IR: More precise indexing
  - Kircz/Nando: selective IR for physics papers (e.g. "maximum entropy" only if in method section)
  - Corston-Oliver: index only clauses with "important" rhetorical relations
- For summarisation: Better content determination
  - Marcu: infer importance from RST tree structure
  - Teufel/Moens: use rhetorical sections with more important propositional content
- For user tailoring in NLG
  - Users of different expertise need different rhetorical information (in a summary, or for within-document navigation)
- For navigation between documents
  - Rhetorical links between web pages; between scientific articles

Rhetorical Structure Theory

- Mann and Thompson, "Rhetorical Structure Theory: A Theory of Text Organisation", ISI/RS-87-190, USC, 1987
- Fixed set of 23 rhetorical relations holding between any two adjacent clauses or larger text segments:

CIRCUMSTANCE	SOLUTION-HOOD	ELABORATION	BACKGROUND
CONTRAST	ENABLEMENT	CAUSE (NON-VOLITIONAL)	JOIN
EVIDENCE	JUSTIFICATION	CAUSE (VOLITIONAL)	SUMMARY
MOTIVATION	CONCESSION	RESULT (NON-VOLITIONAL)	SEQUENCE
PURPOSE	ANTITHESIS	RESULT (VOLITIONAL)	RESTATEMENT
CONDITION	INTERPRETATION	EVALUATION	

- Most relations are asymmetric: nucleus, satellite (subordinate information)
- Relations can apply recursively to non-atomic text pieces
- The analyst provides a plausible reason the writer might have had for including each part of the whole text

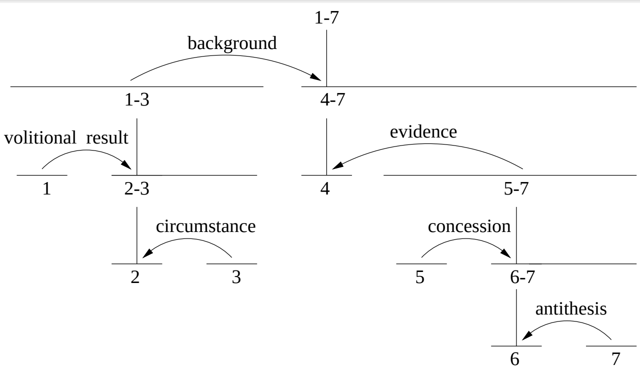
RST: Definition of relations

Relation name: EVIDENCE
Constraints on nucleus: H might not believe Nucleus to a degree satisfactory to S.
Constraints on satellite: H believes Satellite or will find it credible.
Constraints on satellite+nucleus combination: H's comprehending Satellite will increase his believe in Nucleus.
Effect: H's belief in Nucleus is increased.

An EVIDENCE relation with (b) as nucleus:

(a) George Bush supports big business.
(b) He's sure to veto House Bill 1711.

Example: An RST analysis



1 Farmington police had to help control traffic today
2 when hundreds of people lined up to be among the first applying for jobs at the yet-to-open Marriott Hotel.
3 The hotel's help-wanted announcement – for 300 openings – was a rare opportunity for many unemployed.
4 The people waiting in line carried a message, a refutation, of claims that the jobless could be employed if only they showed enough motivation.
5 Every rule has its exceptions,
6 but the tragic and too-common tableaux of hundreds of even thousands of people snake-lining up for any task with a paycheck illustrates a lack of jobs,
7 not laziness.

Practical RST problems

- Many different RST relation inventories exist in literature
- Low human agreement on analyses
- High degree of vagueness during analysis:
  - How should the units of the analysis be determined?
  - At which level in the tree should a given unit connect?
  - Most RST relations are not explicitly marked in text

Theoretical RST Problems

- Moore, Moser (1992, CL): RST analyses are **systematically** ambiguous
- Reason: RST mixes **intentional** and **informational** content

(a) George Bush supports big business.
(b) He's sure to veto House Bill 1711.

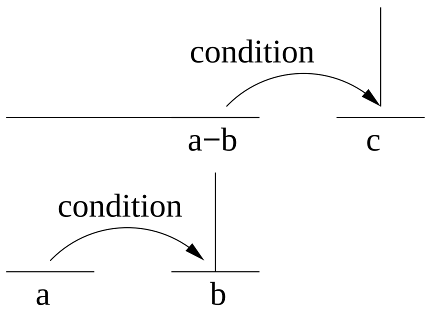
Two Analyses are possible:

- EVIDENCE with nucleus b) (presentational, i.e., intentional relation)
- or VOLITIONAL CAUSE, also with nucleus b) (subject matter, i.e., informational relation)

Moore and Moser, Example

a) Come home by 5:00
b) then we can go to the hardware store before it closes
c) that way we can finish the bookshelves tonight

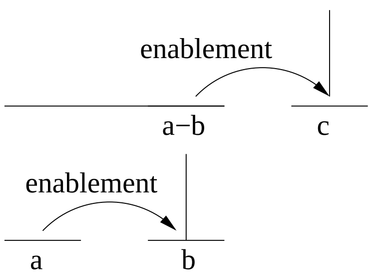
Informational layer:



Moser and Moore, Example

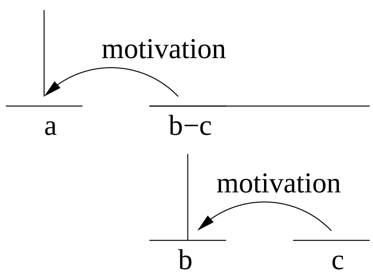
Intentional layer:

A



OR

B



- Informational level alone is not enough
- Informational structure of discourse is different (and not even isomorphic) to intentional structure

A problem for RST: Moore and Moser 92

- Both presentational and subject-matter levels are needed for many practical tasks (e.g. to plan discourse response to answer "It's not necessary to go to the hardware store. I borrowed a saw from Jane.")
  - If intention was (A) to make sure that H realises the shop closes early tonight:
    "OK, I'll come home the ususal time then."
  - If intention was (B), to make H come home at 5:00 (e.g. for a surprise party):
    "Come home by 5:00 anyway or else you'll get caught in the traffic."
- Moore and Moser's RDA (Relational Discourse Analysis) encodes both layers

Robust RST Parsing (Marcu 1997): Algorithm

Algorithm:

1. Identify clause boundaries and discourse markers
2. Determine rhetorical relations
3. Use theorem prover and axioms of correct trees to find all valid trees
4. Choose trees that are skewed to the right

Offline resource: Corpus study of 231 cue phrases (2100 occurrences) and their rhetorical properties

Clause boundary identification

- Mark all potential cue phrases.
- Decide which ones are cue phrases and where the discourse unit boundaries should be:

Marker	Posit.	Action
Although	B	comma
although	B	dual
because	B	dual
but	B	normal
where	B	comma-paren
Yet	B	nothing

Possible actions

nothing	no boundary	*Yet that was not all.*
normal	boundary immediately before cue phrase	*I went home | but left soon afterwards again.*
comma	boundary after next comma, but if comma is followed by *and* or *or*, boundary after next comma (if there is one) or at end otherwise	*Although it was not required, and in fact not even desired, | it did play a role.*
normal-then-comma	before marker and after first comma (in case of encountering an *and* immediately after the comma, delay until next comma, cf. above)	
end	boundary after cue phrase	
match-paren	both at open and closing parenthesis	open parenthesis

comma-paren	before marker and after next comma	*Yet, even on the summer pole, < where the sun remains in the sky all day long, > temperatures are never high enough to melt frozen water.* |
match-dash	before cue phrase (dash) and after matching dash or at end	dash
set-and/set-or	(store info that *and/or* was encountered)	
dual	before marker unless there is other marker immediately before. If there is, do as for comma	*I went to the theatre | although I had a terrible headache.* *I went to the theatre, | and although I had a terrible headache, | I do not regret it.*

A wrong Clause boundary identification

- *I gave John a boat, | which he liked, and a duck, | which he didn't. |*

Recall 80.8% and precision 89.5%, but numerical values inflated as sentence ends (trivial) are counted as correct too

Hypothesise all possible rhetorical relations

Only the midday sun at tropical latitudes is warm enough to thaw ice on occasion [4]| *but any liquid water formed in this way would evaporate almost instantly* [5]| *because of the low atmospheric pressure* [6]|.

Marker	Status	Wh-to-link	Types	Rhet.Rel	Max. dist	Dist sal.
because	S_N	After	Clause	CAUSE, EVIDENCE	1	0
because	**N_S**	**Before**	**Clause**	CAUSE, EVIDENCE	**1**	**0**
but	N_N	Before	Clause	CONTRAST	1	0

Where-to-link: unit containing marker is to be linked to some other unit. Does this unit come BEFORE or AFTER the unit containing marker?

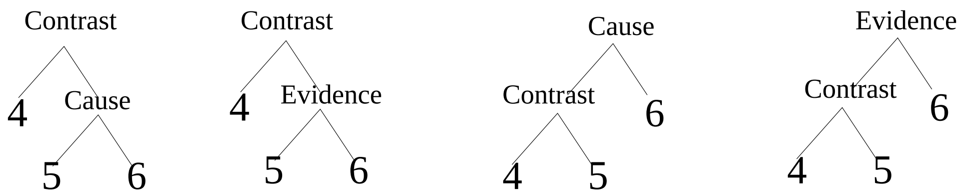
Maximal Distance: maximal number of units of the same kind found between textual units involved in rhetorical relation. 0 means units were always adjacent.

Distance to salient unit: any known cases of rhet. relation holding between

Robust RST Parsing: Relation Hypotheses

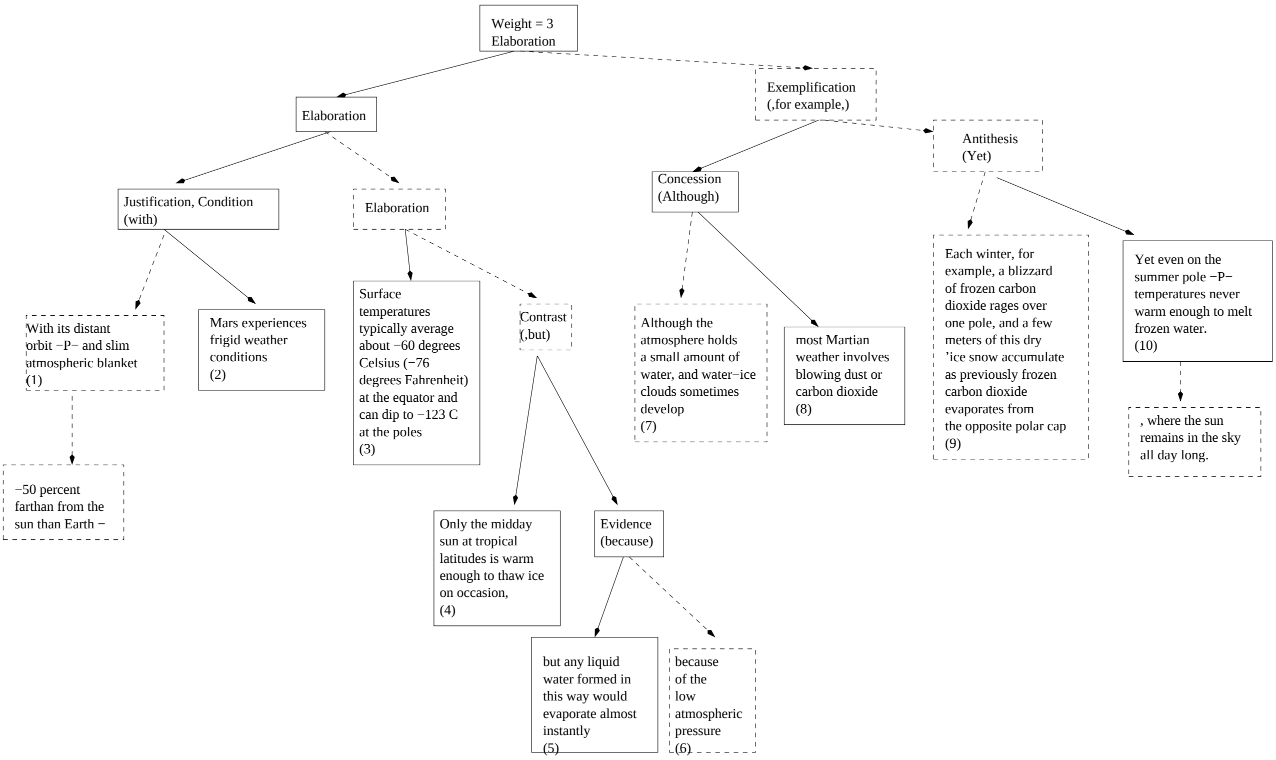
Ambiguities in our example:

- because: 2 rhetorical relations are possible: CAUSE or EVIDENCE
- because: 2 syntactic patterns are possible: "because Y, X" and "X because Y"
- As both *but* and *because* have max. distance 1, all rhetorical relations involved could span from unit 4 to unit 6.



(First tree is correct)

Robust RST Parsing (Marcu 1997): Example output

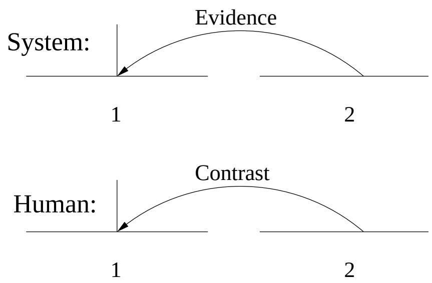


Marcu 1997, Results

	Humans		System	
	R	P	R	P
Units	87.9	87.9	51.2	95.9
Spans	89.6	89.6	63.5	87.7
Nuclearity	79.4	88.2	50.6	85.1
Relations	83.4	83.4	47.0	78.4

Marcu (1997), Discussion

- Unduely raised numerical values



	System	Annotator
1	SPAN	SPAN
2	EVIDENCE	CONTRAST
	1/2 P; 1/2 R	

- System and Annotator do not agree on span → R/P should be 0, but as the empty SPANS are counted it is 50%

Observation 1: Sentiment towards Cited Work

For these reasons numerous Tröger's base derivaties have been prepared . . . [2,3,5]. However, some of the above methodologies possess ***tedious*** *work-up procedures or include relatively* ***strong reaction conditions*** *. . . with* ***poor to moderate yields****, as is the case for analogues* **4** *and* **5**.

→ **Criticised** approach; typically in motivation

The OH BDE values of a series of alkyl- and alkoxy-substituted phenols have been precisely determined by Pedulli and coworkers . . . [24]. This method gives ***accurate*** *BDE values relative to a reference compound, 2,4,6-tri-tert-butyl phenol. We have utilized this experimental data to evaluate the model for BDE determination . . .* (b515712a)

→ **Praised/Used** approach; typically used as part of authors' own solution

Observation 1 Holds Across Disciplines

Previous parser comparisons ... [Tom87, BL89, Sha89, BvN93, MK93]. It is **not clear** *that these results scale up to reflect accurately the behaviour of parsers using realistic, complex unification-based grammars....* (9405033, S-5/6)

→ **Criticised** approach

The technical vehicle previously used to extract the specialized grammar is explanation-based generalization (EBG) [Mit86]. The EBG scheme has previously proved most **successful** *for tuning a natural-language grammar to a specific application domain and thereby achieve* **very much faster** *parsing, at the cost of a small reduction in coverage.* (9405022, S-162/163)

→ **Praised** approach

Citation Context and Sentiment

S−5 Hindle (1990) proposed dealing with the sparseness problem by estimating the likelihood of unseen events from that of "similar" events that have been seen. S−6 For instance, one may estimate the likelihood of a particular direct object for a verb from the likelihoods of that direct object for similar verbs. S−7 This requires a reasonable definition of verb similarity and a similarity estimation method. S−8 In Hindle 's proposal, words are similar if we have strong statistical evidence that they tend to participate in the same events. S−9 His notion of similarity seems to agree with our intuitions in many cases, but it is not clear how it can be used directly to construct word classes and corresponding models of association.

- Dissembling and “meek” citations (MacRoberts and MacRoberts 1986)
- 69% of CONTRAST sentences and 21% of BASIS sentences do not contain the citation itself

Observation 2: Knowledge Claim Zones

Telomeres exist at the ends of eukaryotic chromosomes and can protect the chromo somes... Recently, many G-quadruplex stabilizers have been synthesized and studied for their biological and medicinal activities by many groups. [cit7a] [cit7b] [cit7c] [cit8a] [cit8b] [cit9a] [cit9b]. . . However, few reports of corroles in medicinal or biological applications have been published. [cit11a] [cit11b] [cit11c] [cit11d] In this paper, we shall report our synthesis of cationic corrole derivatives **3** *and* **5** *. . .*

Observation 2: Knowledge Claim Zones

- *Knowledge claim:* New contribution associated with one paper
- Discourse segments can be defined by who owns the knowledge claim:
  - Paper authors ("us")
  - Somebody else ("them")
  - Nobody (future or general)
  - Segments often neutral (with some sentiment around the edges)
  - Cited approaches appear in fixed *roles* or *functions*

Observation 3: Common Sequences of Rhetorical Moves

- Problem, followed by research goal:

> *However, some of the above methodologies possess tedious work-up procedures or include relatively strong reaction conditions, such as treatment of the starting materials for several hours with an ethanolic solution of conc. hydrochloric acid or TFA solution, with poor to moderate yields, as is the case for analogues* **4** *and* **5** *[5]. Considering these potential applications, we now report a simple synthetic method for the preparation of . . .*

Observation 3: Common Sequences of Rhetorical Moves

> *A problem not fully explored yet is how to arrive at an optimal choice of tree-cutting criteria. In the previous scheme, these must be specified manually, and the choice is left to the designer's intuitions. This article addresses the problem of automating this process and presents a method where the nodes to cut at are selected automatically using the information-theoretical concept of entropy.* (9405022, S-17–19)

Observation 4: Argument for Knowledge Claims

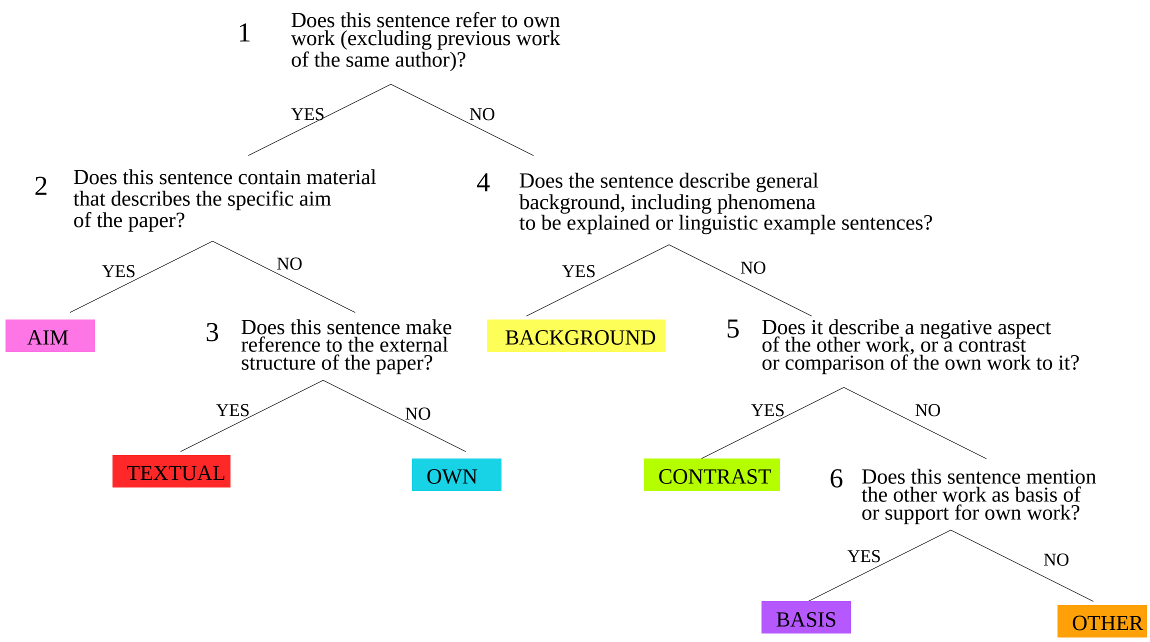
- Scientific Argumentation is fixed and prototypical
- Authors must justify their new knowledge claim
- This provides constraints in a rhetorical game
- They use rhetorical moves to do so

Rules in this game:

- There are two sets of players: "them" and "us"
- There are negative states and positive states (praise, criticism, failed and successful problem-solving)
- Valid knowledge claims move a negative knowledge state to a more positive one

Teufel (2010). The Structure of Scientific Articles: Applications to Citation Indexing and Summarization. CSLI Publications.

Argumentative Zoning



Argumentative Zoning of a Chemistry Paper

Synthesis of pyrazole and pyrimidine Troeger's base–analogues

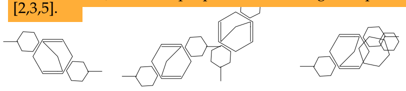
1 PERKIN

Rodrigo Abonia, Andrea Albernez, Hector Larrabondo, Jairo Quiroga, Braulio Isuasty, Henry Isuasty, Angelina Hormaza Adolfo Sanchez, and Manuel Nogueras

Troeger's–base analogues bearing fused pyrazolic or pyrimidinic rings were prepared in acceptable to good yields through the reaction of 3–alkyl–5–amino–1arylyrazoles and 6–aminopyrimidin–4(3H)–ones with formaldehyde under mild conditions (i.e. in ethanol at 50C in the presence of catalytic amounts of acetic acid. Two key intermediates were isolated from the reaction mixtures, which helped us to suggest a sequence of steps for the formation of the Troeger's bases obtained. The structures of the products were assigned by 1H and 13 CNMR, mass spectra and elemental analysis and confirmed by X–ray diffraction for one of the obtained compounds.

Introduction

Although the first Troeger's base 1 was obtained more than a century ago from the raction of p–toluidine and formaldehyde [11], recently the study of these compounds has gained importance due to their potential applications. They possess a relatively rigid chiral structure which makes them suitable for the development of possible synthetic enzyme and artificial receptor systems [2], chelating and biomimetic systems [3] and transition metal complexes for regio–and stereoselective catalytic reac–tions [4]. For these reasons, numerous Troeger's–base derivates have been prepared bearing different types of substituents and structures (i.e. 2–5 Scheme 1), with the purpose of increasing their potential applications [2,3,5].



Scheme 1 The original Troeger's–base 1 and some interesting deri–vatives and analogues.

However, some of the above methodologies possess tedious work–up procedures or include relatively strong reaction conditions, such as treatment of the starting materials for several hours with an ethanolic solution of conc. hydrochloric acid or TFA solution, with poor to moderate yields, as is the case for analogues 4 and 5 [5].

Considering these potential applications, we now report a simple synthetic method for the preparation of 5,12–dialkyl–3,10–diaryl–1,3,4,8,10,11–hexa–azetetracyclo[6.6.1.0 2,6 .0 9,13] pentadeca–2(6),4,9(13),11–tetraenes 8a–e and 4,12–dimethoxy–1,3,5,9,11,13–hezaaatetrctyclo[7.7.1.0 2,7.010,15] heptadeca2(7),3,10(15)m11–tetraene–6m14–diones 10a,b based on thereaction of 3–alkyl–5–amino–1–arylpyrazoles 6 and 6–aminopyrimin–4(3H)–ones 9 with formaldehyde in ethanol and catalytic amounts of acetic acid. Compounds 8 and 10 are new Troegers base analogues bearing heterocyclic rings instead of the usual phenyl rings in their aromatic parts.

Results and discussion

In an attempt to prepare the benzotriazolyl derivative 7a, which could be used as in intermediate in the synthesis of new hydroquinolines of interest, [6], a mixture of 5–amino–3–methy–1–phenylpyrazole 6a,formaldehyde and benz,otri–azole in 10 ml of ethanol , with catalytic amounts of acetic acid, weas heated at 50C for 5 minutes. A solid precipidated from the solution while it was still hot. However, no consumption of benzotriazole was observed at TLC.

The reaction conditions were modified and the same product was obtained when the reaction was carried out without using benzotriazoole, as shown in Schema 12. On the basis of NMR and mass spectra and X–ray crystallographic analysis we established that the structure is 5,12–diakyl–3 10–diaryl–1,3,4,8,10,11–hexa

pentacyclic Troeger's base analogue.

Co_Gro	Other	Aim	Gap/Weak	Own_Mthd	Own_Res	Own_Conc

Grosz and Sidner (1986)
RST
Marcu 's (1997) RST Algorithm
KCDM – Knowledge Claim Discourse Model

Argumentative Zoning of a Computational Linguistics Paper

Distributional Clustering of English Words

Fernando Pereira Naftali Tishby Lillian Lee

Abstract

We describe and experimentally evaluate a method for automatically clustering words according to their distribution in particular syntactic contexts. Deterministic annealing is used to find lowest distortion sets of clusters. As the annealing parameter increases, existing clusters become unstable an subdivide, yielding a hierarchical "soft" clustering of the data. Clusters are used as the basis for class models of word occurrence, and the models evaluated with respect to held–out test data.

Introduction

Methods for automatically classifying words according to their contexts of use have both scientific and practical interest. The scientific questions arise in connection to distribution–al views of linguistic (particularly lexical) structure and also in relation to teh question of lexical acquisition both from psychological and computational learning perspectives. From the practical point of view, word classification addresses questions of data sparseness and generalization in statistical language models, particularly models for deciding among alternatives analyses proposed by a grammar.

It is well–known that a simple tabulation of frequencies of certain words participating in certain configurations, for example the frequencies of pairs of a transitive main verb and the head noun of its direct object, connot be reliably used for comparing the likelihoods of different alternative configurations. The problem is that for large enough corpora the number of joint events is much larger than the number of event occurrences in the corpus, so many events are seen rerely or never, making their frequency counts unreliable estimates of their probabilities.

Hindle (1999) proposed dealing with the sparseness problem by estimating the likelihood of unseen events from that of "similar" events that have been seen. For instance, one may estimate the likelihood of a particular direct object of a verb from the likelihoods of that direct object for similar verbs. This requires a reasonable definition of verb similarity and a similarity estimation method. In Hindle's proposal, words are similar if we have strong statistical evidence that they tend to participate in the same events. His notion of similarity seems to agree with our intuitions in many cases, but it is not clear how it can be used directly to construct word classes and corresponding models of association.

Our research addresses some of the same questions and uses similar raw data, but we investigate how to factor word association tendencies into associations of words to certain hidden sense classes and associations between the classes themselves. While it may be worthwhile to base such a model on preexisting classes (Resnik, 1992), in the work described here we look at how to derive the classes directly from distributional data. More specifically, we model senses as probabilistic concepts or clusters c with corresponding cluster membership probabilities p(c|w) for each word w. Most other class–based modeling techniques for natural language rely on "hard" Boolean classes (Brown et al., 1990). Class construction is then combinatorically very demanding and depends on frequency counts for joint events involving particular words, a potentially unreliable source of information as we noted above. Our approach avoids both problems.

Problem Setting

In what follows, we will consider two major word classes, V and N, for the verbs and nouns in our experiments, and a single relation between them, in our experiments the relation between a transitive main verbs and the head noun of its direct object. Our raw knowledge about the relation consists of the frequencies fvn of occurrence of particular pairs (v, n) in the required configuration in our corpus. Some form of text analysis is required to collect such a collection of pairs. The corpus used in our first experiment was derived from newswire text automatically parsed by Hindle's parser Fiddich (Hindle, 1993). More recently, we have constructed similar tables with the help of a statistical part–of–speech tagger (Church, 1988) and of tools for regular expression pattern matching on tagged corpora (Yarowsky, 1992). We have not yet compared the accuracy and coverage of the two methods, or what systematic biases they might introduce, although we took care to filter out certain systematic errors, for instance the mis–parsing of the subject of a complement clause as the direct object of a main verb for report verbs like "say".

We will consider here only the problem of classifying nouns according to their distribution as direct objects of verbs; the converse problem is formally similar. More generally, the theoretical bias for our methods supports the use of clustering to build models for any n–ary

Features for Recognition

Type	Name	Feature description	Values
Absolute Location	`Loc`	Position of sentence in relation to 10 segments	10
Explicit Structure	`Section Struct`	Relative and absolute position of sentence within section	7
	`Para Struct`	Relative position of sentence within a paragraph	3
	`Headline`	Type of headline of current section	16
Sentence length	`Length`	Sentence longer than 12 tokens?	2
Content Features	`Title`	Does the sentence contain words from the title or headlines?	2
	`TF*IDF`	Does the sentence contain “significant TFIDF terms”?	2
Verb Syntax	`Voice`	Voice (of first finite verb in sentence)	3
	`Tense`	Tense (of first finite verb in sentence)	10
	`Modal`	Is the first finite verb modified by modal auxiliary?	3
Citations	`Cit`	Citation present? Self citation? Location of citation?	10
History	`History`	Most probable previous category	8
Meta-discourse	`Formulaic`	Type of formulaic expression occurring in sentence	28
	`Agent`	Type of Agent	10
	`Action`	Type of Action, with or without Negation	28

Action Types

Action Type	Example
AFFECT	*we hope to improve our results*
ARGUMENTATION	*we argue against a model of*
AWARENESS	*we are not aware of attempts*
BETTER_SOLUTION	*our system outperforms . . .*
CHANGE	*we extend CITE's algorithm*
COMPARISON	*we tested our system against. . .*
CONTINUATION	*we follow Sag (1976) . . .*
CONTRAST	*our approach differs from . . .*
FUTURE_INTEREST	*we intend to improve . . .*
INTEREST	*we are concerned with . . .*
NEED	*this approach, however, lacks. . .*
PRESENTATION	*we present here a method for. . .*
PROBLEM	*this approach fails. . .*
RESEARCH	*we collected our data from. . .*
SIMILAR	*our approach resembles that of*
SOLUTION	*we solve this problem by. . .*
TEXTSTRUCTURE	*the paper is organized. . .*
USE	*we employ Suzuki's method. . .*
COPULA	*our goal is to. . .*
POSSESSION	*we have three goals. . .*

Entity Types (for "US")

(we/I)
(we/I) also
(we/I) now
(we/I) here
(our/my) JJ* (account/ algorithm/ analysis/ analyses/ approach/ application/ architecture. . .)
(our/my) JJ* (article/ draft/ paper/ project/ report/ study)
(our/my) JJ* (assumption/ hypothesis/ hypotheses/ claim/ conclusion/ opinion/ view)
(our/my) JJ* (answer/ accomplishment/ achievement/ advantage/ benefit. . .)
(account/ . . .) (noted/ mentioned/ addressed/ illustrated . . .) (here/below)
(answer/ . . .) given (here/below)
(answer/ . . .) given in this (article/ . . .)
(first/second/third) author
one of us
one of the authors

- Anaphora resolution for ambiguous NPs (Siddharthan, Teufel, 2007)
- Lexical acquisition of meta-discourse (Abdalla, Teufel 2006).

Performance per category

AZ (2007): $\kappa = 0.48$; Macro-F $= 0.54$

	AIM	TEXT	BKG	OTH	OWN	BASIS	CTR
P	0.59	0.60	0.48	0.59	0.83	0.50	0.46
R	0.63	0.66	0.46	0.40	0.92	0.30	0.31
F	0.61	0.63	0.47	0.48	0.87	0.38	0.37

Citation map: a paper

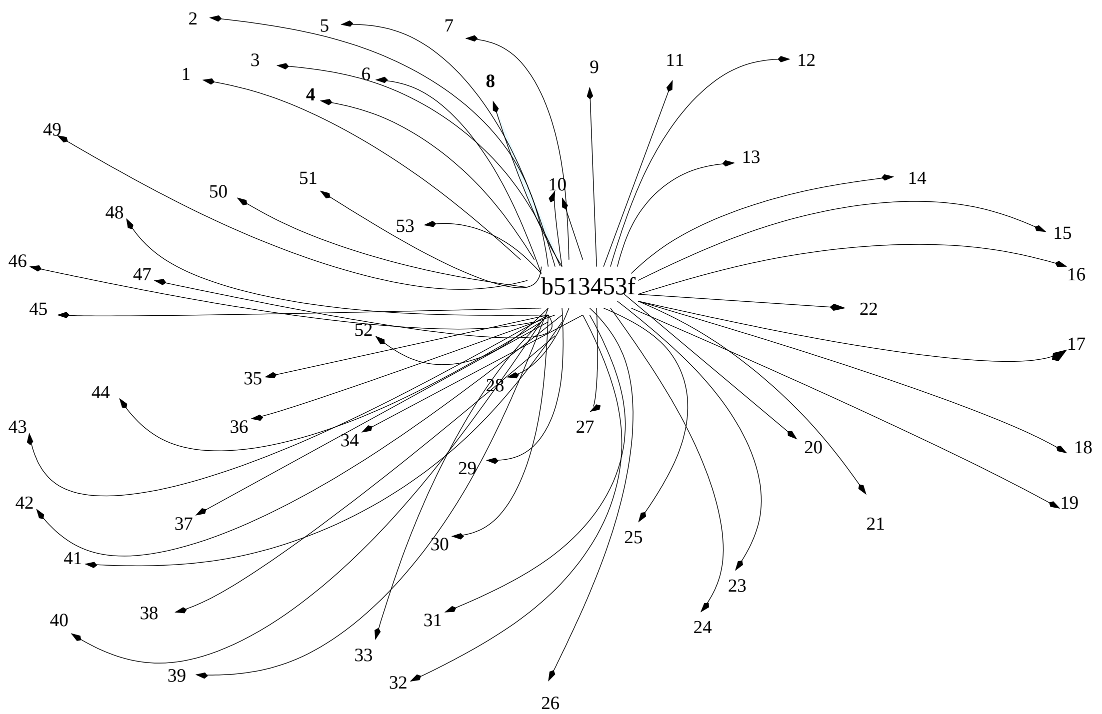
b513453f

Deetlefs, Seddon, Shara
Phys. Chem. Chem Phys.
2006, 8, 642–649

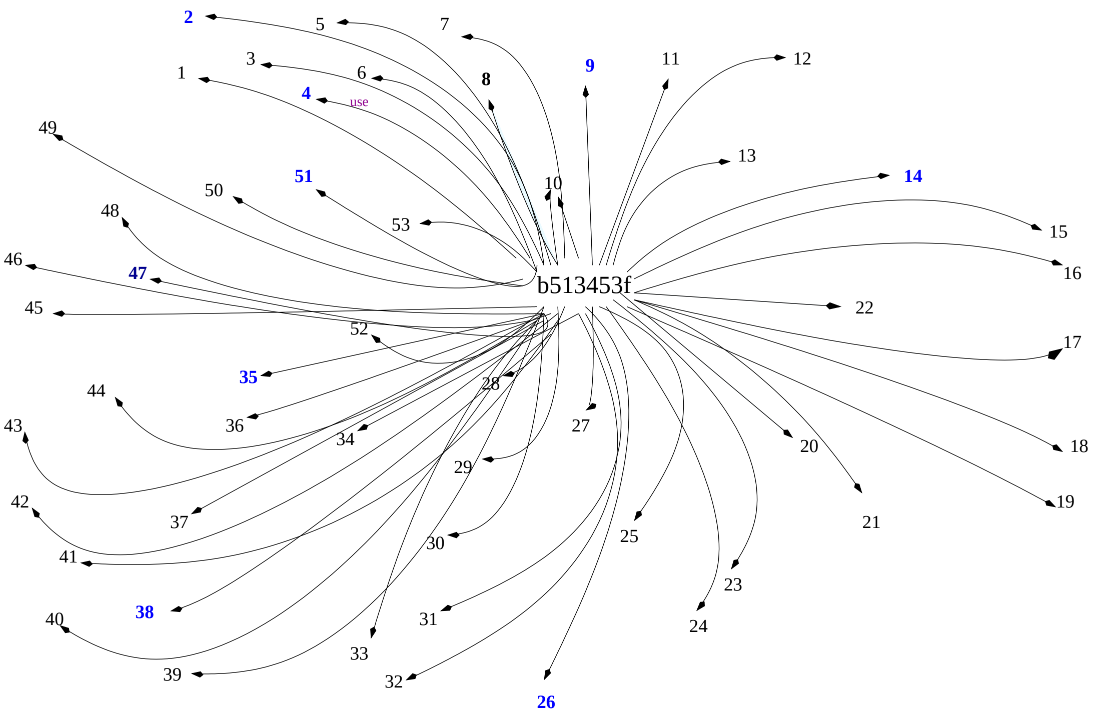
"Predicting Physical Properties of ionic liquids"

A simple method to predict the densities of a range of ionic liquids from their surface tensions, and vice versa, using a surface–tension–weighted molar volume, the parachor, is presented.

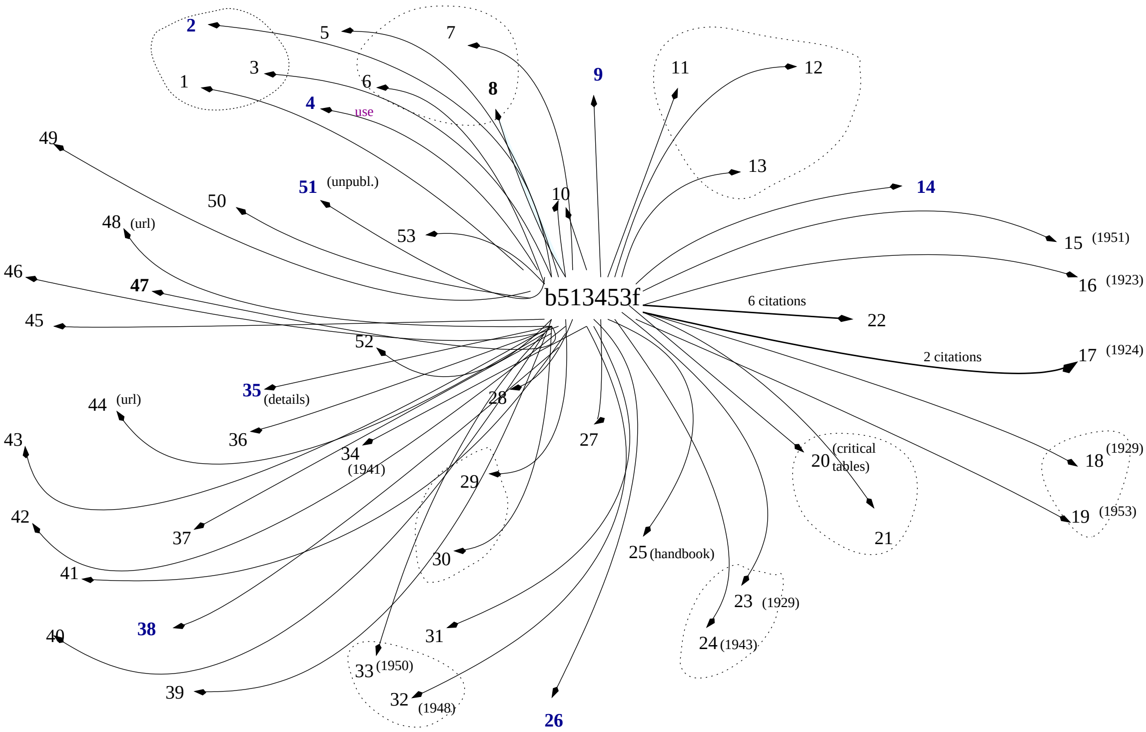
Citation map: all 53 cited papers



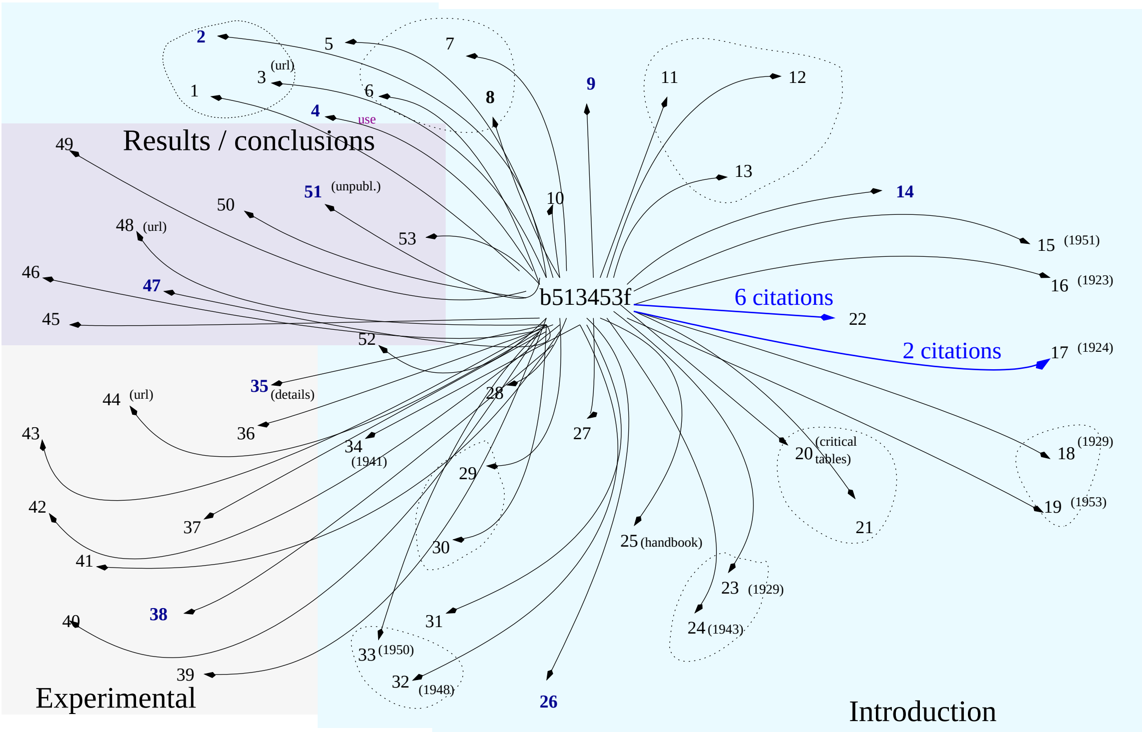
Citation map: self-citations



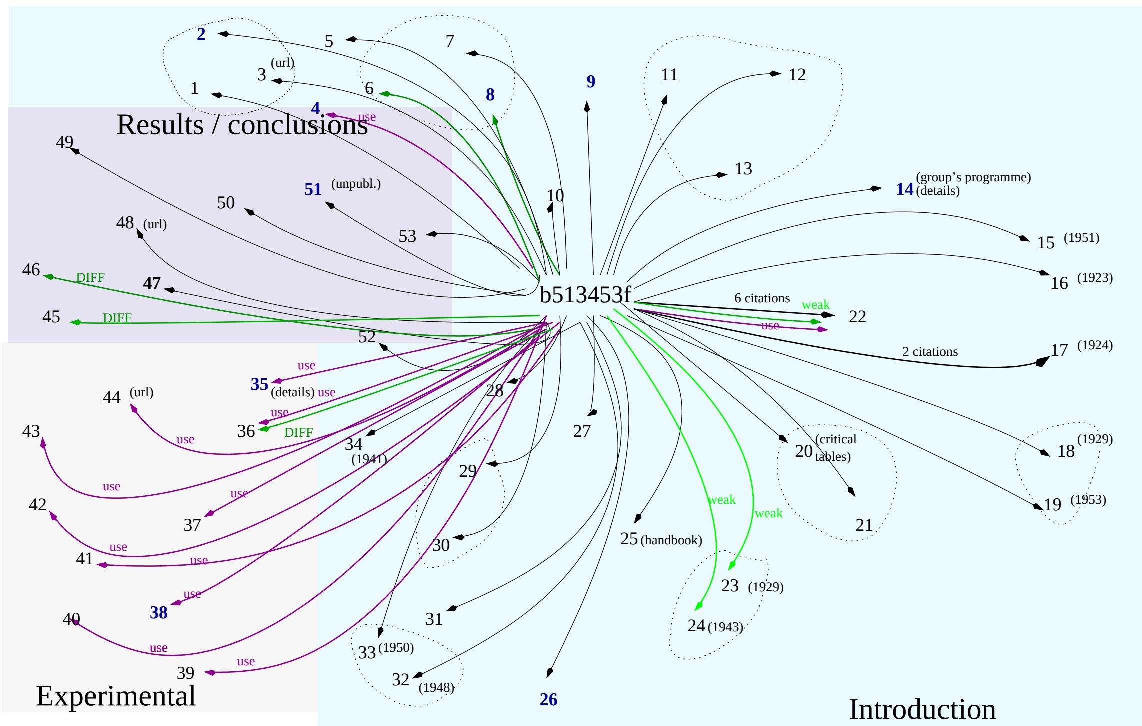
Citation map: number of citations per sentence



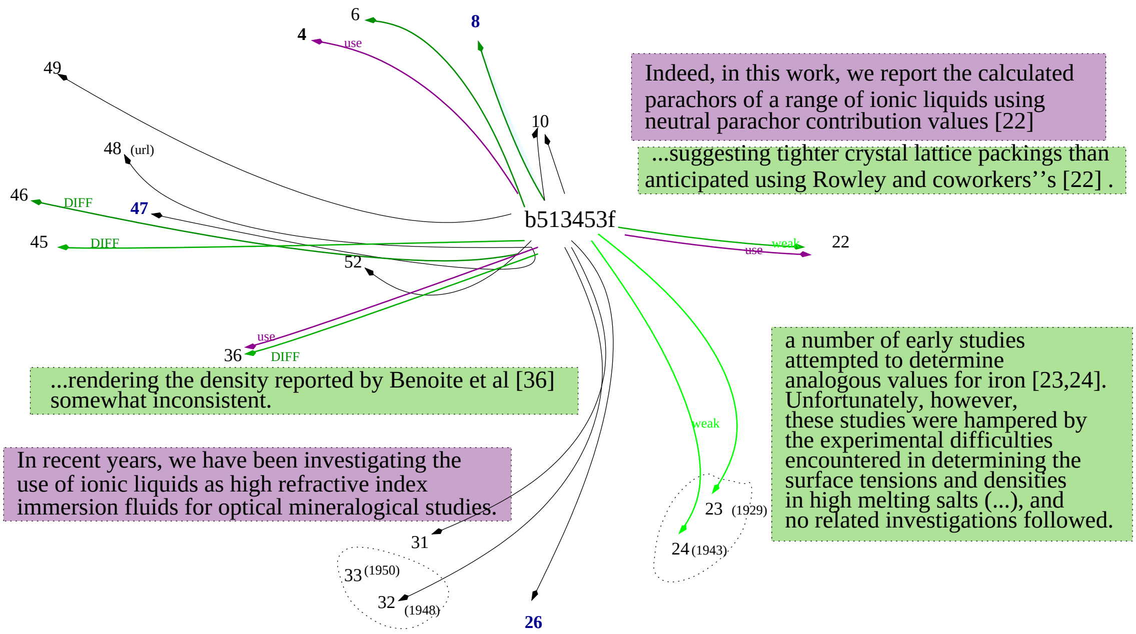
Citation map: rhetorical sections + # citation instances



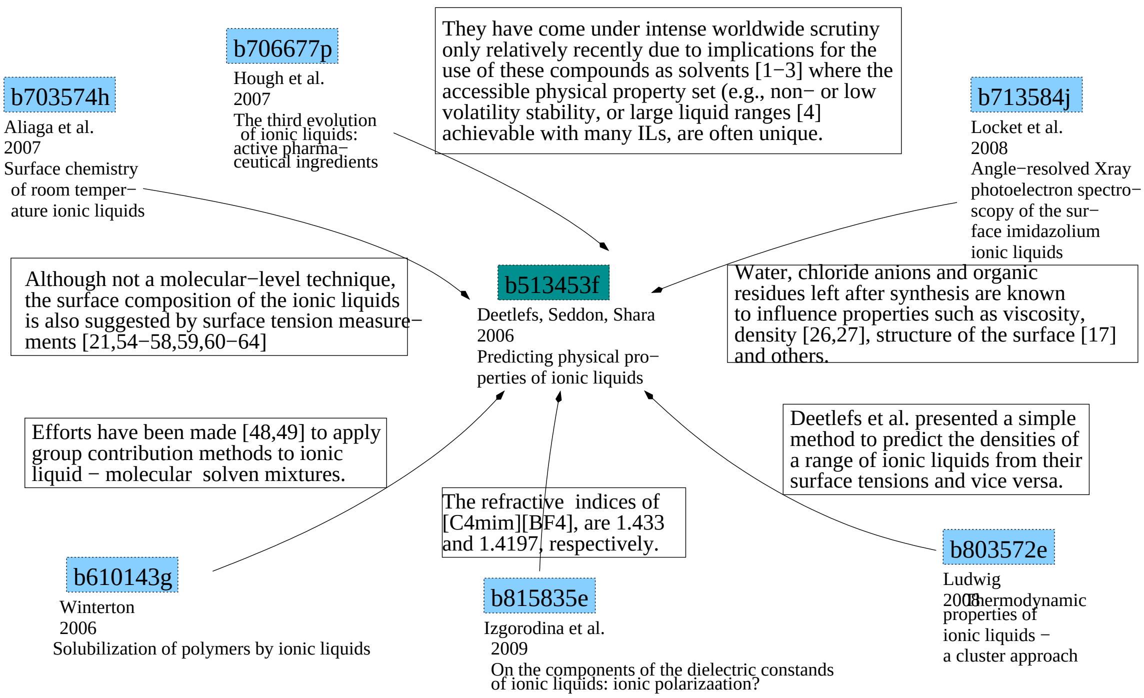
Citation map: sentiment/citation function



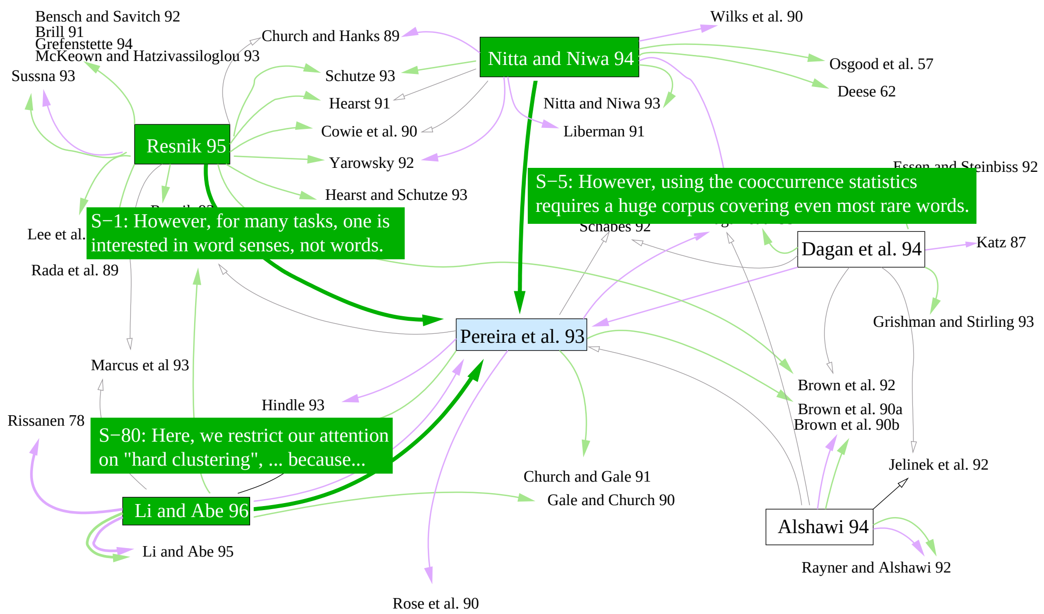
Citation map: connections between papers (outgoing cits.)



Citation map: connections between papers (incoming cits.)



Citation map in Computational Linguistics



Another application: Teaching of argumentation in science

- Recognise rhetorical faux-pas of unpracticed writers in science
- Suggest alternative meta-discourse, ordering etc.
- Feltrim et al. (2005) ported AZ Features to Portugese
- Tool critiques students' introductions of Brazilian CS theses

Feltrim, Teufel, Gracas Nunes and Alusio (2005). Argumentative Zoning applied to Critiquing Novices' Scientific Abstracts In *Computing Attitude and Affect in Text: Theory and Applications* Shanahan, J.; Qu, Y.; Wiebe, J. (Eds.) 2005, Spinger.

Robust AZ

- Collaboration with Min-Yen Kan from National University of Singapore
  - Goal: perform AZ when input is not in SciXML (e.g. ACL anthology via pdfbox)
  - Features used: all words in sentence, citations, history by second-pass, location, presence of cue word
  - Try it out at:

`http:www.wing.nus.edu.sg/zoning`

Conclusion

- Grosz and Sidner 86: Focus space, intentional structure and linguistic effects co-constrain each other
- Too much world knowledge necessary for the general case to implement this model
- RST: General model, but relations cover only small local texts (1-2 paragraphs)
- Implementation possible, but human agreement somewhat low
- Theoretical problems
- KCDM: Rhetorics/argumentation in science
- Exploit easy recognisability of certain moves (e.g., novelty claim) and sentiment to deduce others
- Machine-learn indicators
- Reasonable results, but meta-discourse across disciplines varies → model needs retraining
- Many applications: bibliometrics, search, authoring support