

Modern and Historical Aspects of the UCREL Semantic Analysis System

Paul Rayson, Scott Piao, Dawn Archer

UCREL

Department of Computing,

Department of Linguistics and English Language

Lancaster University, UK.





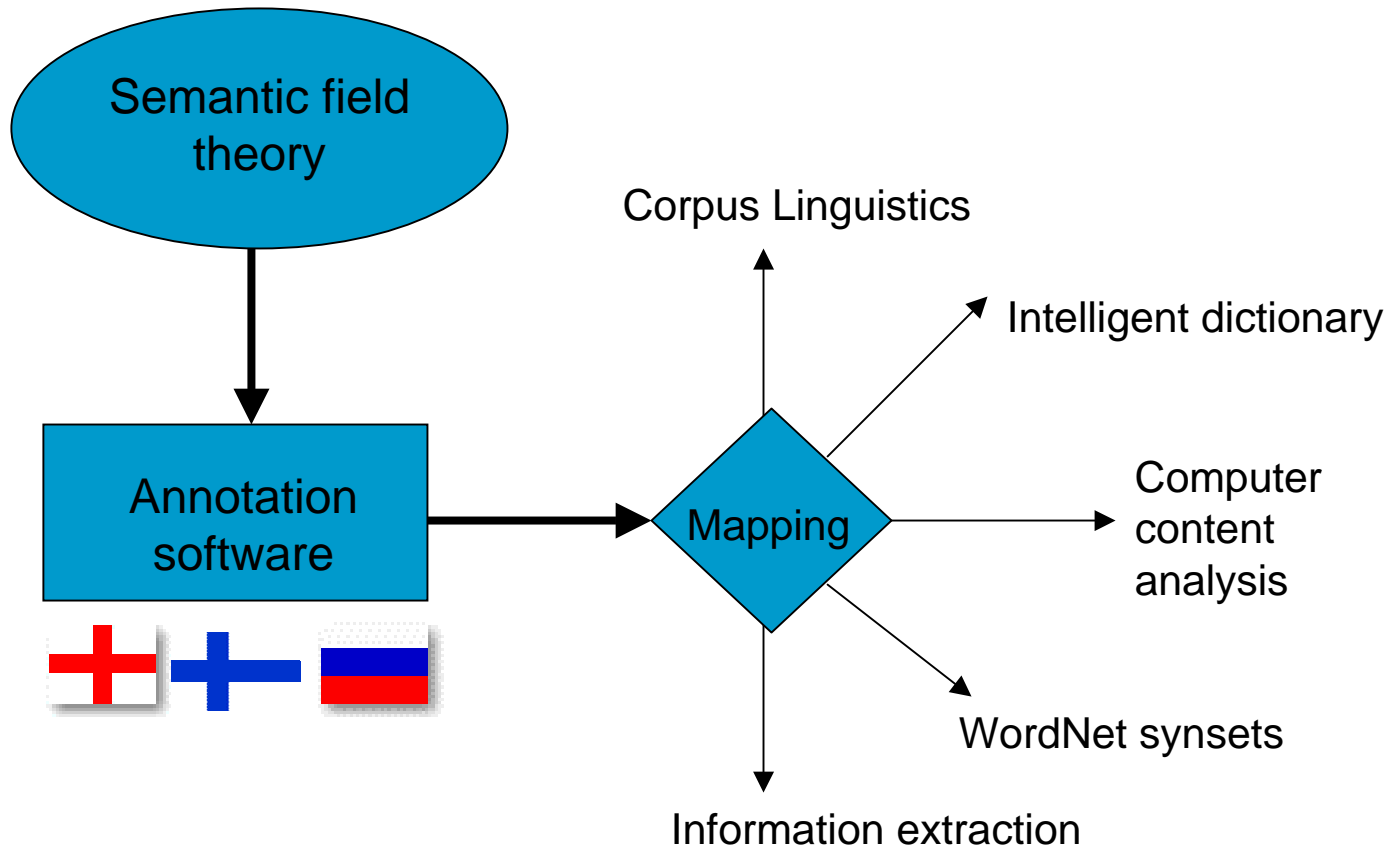
Outline

- The purpose of this talk is to introduce the UCREL Semantic Analysis System (henceforth USAS), an automated software tool for the semantic field annotation of running texts. The talk will consist of three parts:
 - How and why USAS was developed.
 - Its recent application as part of the Benedict project, including its porting to the Finnish language.
 - The modifications we have made to the system so that it can cope with historical texts ranging from the 1600s onwards.
- For more information, see: www.comp.lancs.ac.uk/ucrel/usas/
- Acknowledgements:
 - Original PIs were Roger Garside, Geoffrey Leech, Jenny Thomas
 - Currently funded EU IST Benedict project IST-2001-34237

How and why USAS was developed

Paul's bit

Overview





Application contexts

- Semantic field analysis
- Content analysis
 - Conceptual analysis: USAS, Louw/Nida categories in OpenText.org
 - General category: General Inquirer, Minnesota Contextual Content Analysis
 - Specialised content analysis: RID, Diction
- Market research interview transcript analysis
- Word sense disambiguation: Senseval
- Information extraction / text mining
- Electronic dictionaries



Information extraction

- Requirements reverse engineering to support business process change (Revere)
- Reducing rework through decision management (Tracker)

Links to Lexicography

- The New Intelligent Dictionary (Benedict)
- Providing an interactive user-specified access interface, tailoring the dictionary information supply according to user specifications, incorporating multi-layered entry structure with new information categories and links to corpus data and syntactically- and semantically-based corpus search tools in the dictionary data base.





The task we set ourselves

- Full text tagging, not just selected words
- Tagging the sense in context, not just the word
- Not task specific categories
- Tag set should make sense (psycho)- linguistically
- Flexible category set with hierarchical structure
- Words and multi-word expressions e.g. phrasal verbs (*stubbed out*), noun phrases (*riding boots*), proper names (*United States of America*), true idioms (*living the life of Riley*)



Semantic fields

- AKA conceptual field, a semantic domain, a lexical field, or a lexical domain
- ‘groups together word senses that are related by virtue of their being connected at some level of generality with the same mental concept’
- Not only *synonymy and antonymy* but also *hyponymy and hypernymy*
- E.g. EDUCATION: academic, coaching, coursework, deputy head, exams, PhD, playschool, revision notes, studious, swot, viva

The UCREL Semantic Analysis System

- Hierarchy of 21 major discourse fields expanding into 232 category labels:

Table 1 : The top level of the USAS system

A: General & Abstract Terms	B: The Body & the Individual	C: Arts & Crafts	E: Emotional Actions, States & Processes
F: Food & Farming	G: Government & the Public Domain	H: Architecture, Building Houses & the Home	I: Money & Commerce in Industry
K: Entertainment, Sports & Games	L: Life & Living Things	M: Movement, Location, Travel & Transport	N: Numbers & Measurement
O: Substances, Materials, Objects & Equipment	P: Education	Q: Linguistic Actions, States & Processes	S: Social Actions, States & Processes
T: Time	W: The World & Our Environment	X: Psychological Actions, States & Processes	Y: Science & Technology
Z: Names & Grammatical Words			

Lexical resources

- Lexicon of 44,668 items
 - workshop NN1 I4/H1 P1
- MWE list of 18,553 items
 - travel_NN1 card*_NN* M3/Q1.2
- A small wildcard lexicon
 - *kg NNU N3.5
- A small context rule set of 350 items
 - VB*[Z5] (R*n) (XX) (R*n) V*G*
- Unknown words using WordNet synonym lookup



Disambiguation methods (1)

■ 1. POS tag

- *spring* temporal noun [season sense]
- *spring* common noun [coil sense] [water source sense]
- *spring* verb [jump sense]

■ 2. General likelihood ranking for single-word and MWE tags

- *green* referring to [colour] is generally more frequent than *green* meaning [inexperienced]

■ 3. Overlapping MWE resolution

- Heuristics applied: semantic MWEs override single word tagging, length and span of MWE also significant



Disambiguation methods (2)

- 4. Domain of discourse
 - adjective *battered*
 - [Violence] (e.g. battered wife)
 - [Judgement of Appearance] (e.g. battered car)
 - [Food] (e.g. battered cod)
- 5. Text-based disambiguation
 - one sense per text
- 6. Context rules
 - *Auxiliary verbs (be/do/have)*
 - *account* of NP [narrative]
 - balance of xxx *account* [financial]



Disambiguation methods (3)

■ 7. Local probabilistic

- *account* occurring in the company of *financial, bank, overdrawn, money*
- surrounding words, POS tags or semantic fields
- span of words
- co-occurrence measures rather than HMM

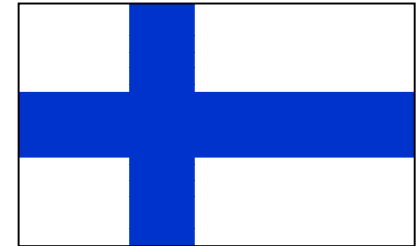
Its recent application as part of the Benedict project, including its porting to the Finnish language

A decorative horizontal bar consisting of a series of colored segments in shades of blue, teal, yellow, and black, arranged in a slightly wavy pattern across the width of the slide.

Scott's bit

USAS in the Benedict Project

- Restructuring and updating of USAS
- Integrating USAS into MT System
- Porting USAS to Finnish language
- Future possibilities



The Benedict Project

- EC Project (ref. IST-2001-34237)
- **Aim:** develop a new methodology and a prototype of intelligent machine dictionary.
- **Project partners:** Lancaster University, HarperCollins Publisher, Kielikone Ltd., University of Tampere, Gummerus Kustannus Ltd and Nokia
- **Websites:**
<http://www.comp.lancs.ac.uk/ucrel/projects.html#benedict>
<http://mot.kielikone.fi/benedict/>



Restructuring and Updating USAS (1)

- Why the trouble?
 - The original tagger is designed in procedural approach, difficult to interact with/integrate into other program/systems.
 - The original tagger is C-coded and difficult to be ported across different platforms (from UNIX to Windows in our case).
 - Difficult to port it to other languages.



Restructuring and Updating USAS (2)

What has been done?

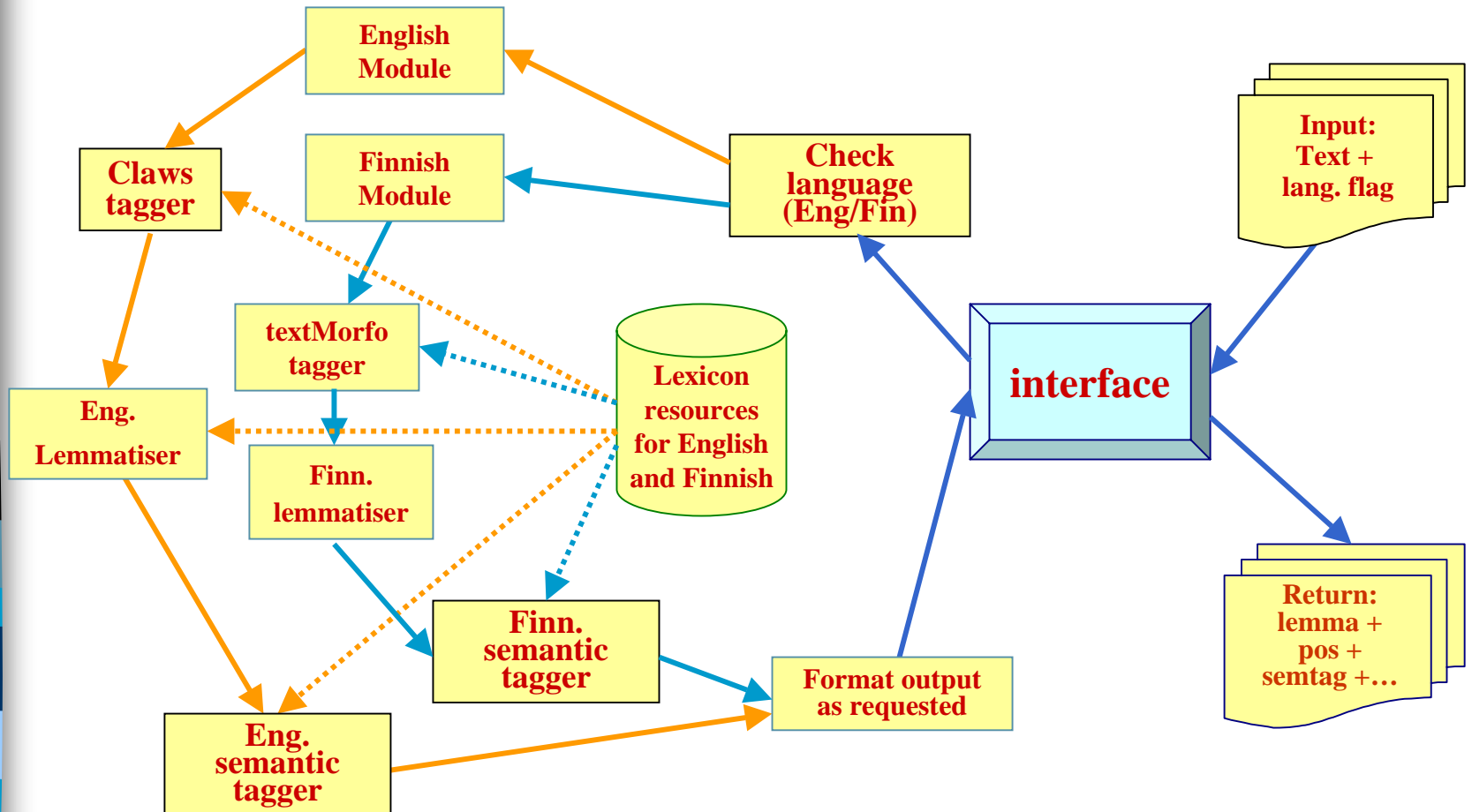
- The original package has been re-structured into an Object-oriented (OO) architecture.
- Semantic tagger has been re-coded in Java; a Windows version of the CLAWS tagger (C-code) is integrated using JNI.
- It Interacts with other Benedict programs through an interface.
- Improved maintainability – each component can be updated/modified without affecting other components.
- The Unicode-based text processing – may be not so useful for English, but can be critical for some other languages.



Porting USAS to Finnish Language

- Building Finnish semantic tagger (FST) is one of the aims of the Benedict Project.
- FST is based on the architecture of the English tagger with only minor modification (mainly for different POS tagger).
- A Finnish morphological analyser *textMorfo*, provided by Kielikone Oy, is used in the place of CLAWS tagger.
- A beta version of Finnish semantic lexical resources are built and provided by Finnish project partners.
- Finnish semantic tagger has become a module of the USAS package.

Structure of USAS (v2)





Evaluation of English Semantic Tagger

- The English semantic tagger has been evaluated on various test data.
- Difficulty experienced: Very few manually semantically tagged data are available.
- Our evaluation:
 - Hand tagged test corpus
 - MWE (Multiword Expression) recognition
 - Lexical coverage



Evaluation on Hand-tagged Corpus (1)

- Test corpus size = 124,839 words
- Error rate = 8.95%
- Ambiguity ratio
 - *Before disambiguation:*
Ambiguity ratio = 47.73%
 - *After disambiguation:*
Ambiguity ratio = 17.06%

Evaluation on Hand-tagged Corpus (2)

Breakdown of tagging methods

Tagging method	Freq. %
Lexicon	63.68
Lexicon after suffix strip	3.41
Lexicon on lemma	0.03
Auto-tag rule	0.39
Domain of discourse	7.67
Auxiliary verb	6.76
Context rules	0.83
Lexicon ignoring POS	0.92
Lexicon stem ignoring POS	0.07
WordNet	0.05
Wildcard MWE	0.54
Normal MWE	11.60
MWE and domain of discourse	4.06

Evaluation on Hand-tagged Corpus (3)

Breakdown of Errors by POS

POS tag	% Error relative to testbed	% Error relative to tag
A:Article	0.21	2.47
C:Conjunction	0.05	0.60
D:Determiner	0.21	4.69
E:Existential there	0.01	1.22
F:Formula Foreign word	0.00	0.31
G:Genitive	0.01	6.62
I:Preposition	0.36	4.16
J:Adjective	0.87	17.65

M:number	0.29	23.93
N:noun	2.62	16.29
P:pronoun	0.06	0.51
R:adverb	1.08	13.47
T:to	0.11	7.52
U:interjection	0.02	0.94
V:verb	3.03	13.21
X:not	0.01	1.25
Z:letter	0.00	2.67
TOTAL	8.95	

Evaluation on MWE Recognition

MWE length	Total MWEs	Accepted MWEs	Precision
2	3,378	3,105	91.92%
3	700	575	82.14%
4	95	91	95.44%
5	18	17	94.44%
6	4	4	100.00%
Total	4,195	3,792	90.39%

For further details, see: Scott Songlin Piao, Paul Rayson, Dawn Archer and Tony McEnery (2005). Comparing and Combining A Semantic Tagger and A Statistical Tool for MWE Extraction. Computer Speech and Language. (to appear).

Evaluation on Lexical Coverage (1)

Modern English Corpora

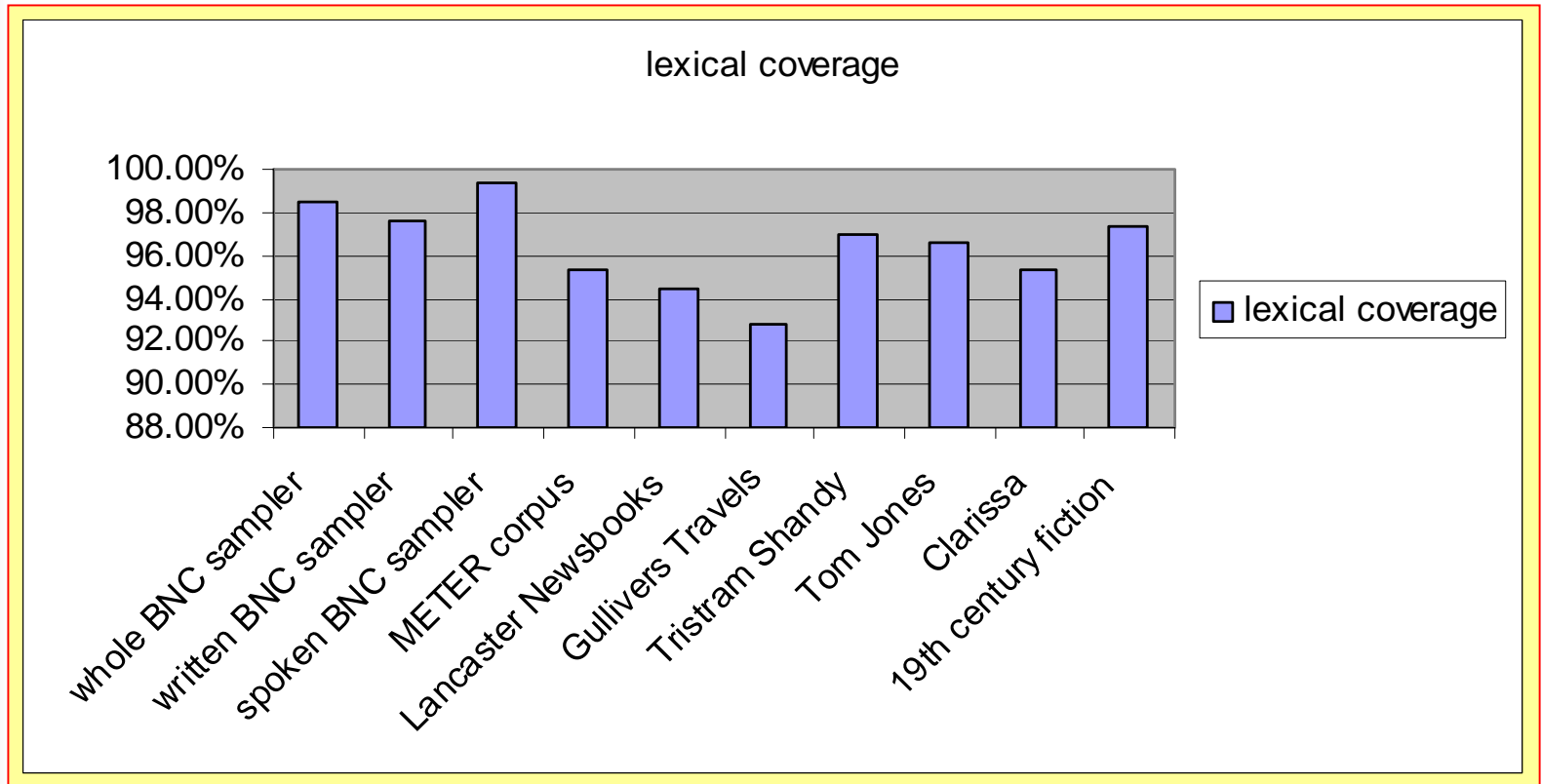
Test Corpus	Total Tokens	Unmatched Tokens	Lexical Coverage
Whole BNC Sampler	1,956,171	29,517	98.49%
BNC Samp. Written sect.	970,532	23,407	97.59%
BNC Samp. Spoken sect.	985,639	6,110	99.39%
METER Corpus	241,311	11,143	95.38%

Evaluation on Lexical Coverage (2)

Historical English Corpora

Test Corpus	Total Tokens	Unmatched Tokens	Lexical Coverage
Lancaster Newsbooks	61,065	3,418	94.40%
Gulliver's Travels	194,987	14,117	92.76%
Tristram Shandy	108,137	3,235	97.01%
Tom Jones	352,942	11,944	96.62%
Clarissa	887,276	40,988	95.38%
19th century fiction	5,000,000	135,661	97.29%

Evaluation on Lexical Coverage (3)



For further details, see: Scott S. L. Piao, Paul Rayson, Dawn Archer and Tony McEnery (2004) Evaluating lexical resources for a semantic tagger. In Proceedings of LREC-04. Lisbon, Portugal. pp. 499-502.



Development of an historical tagger

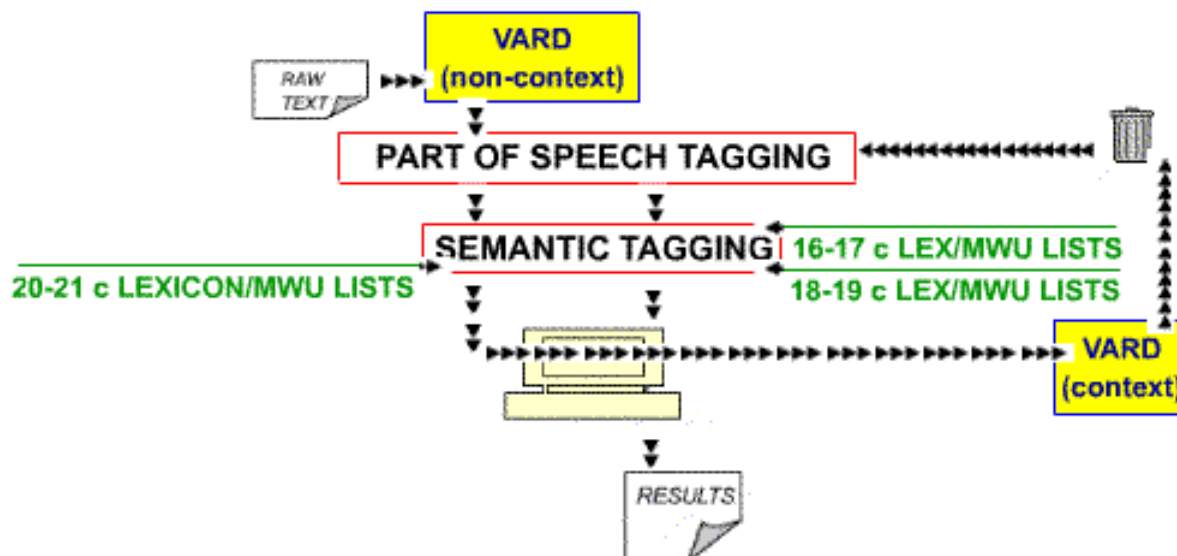
Dawn's bit



Problems and solutions

- Variant spellings
e.g. *bee*, *doe*
-(e)th, *-(e)st*,
(e)s genitive
- Change in meaning
- VARD
 - Non-contextualised
(search and replace script)
 - Contextualised
(context rules)
- Different lexicons for different periods
 - 16th and 17th C
 - 18th and 19th C
 - 20 and 21st C

Historical Tagger



Resources for 16-17th lexicon:

Nameless Shakespeare, Lampeter Corpus (1640-1740), Lancaster Newsbooks (1653/54), Corpus of English Dialogues (1560-1760)

Resources for 18-19th lexicon:

18th and 19th fiction from ProQuest (EEBO texts)



Example VARDED text: Midsummer Night's Dream

Theseus. Now <reg o="faire"> fair Hippolita, our <reg o="nuptiall"> nuptial <reg o="houre"> hour <reg o="Drawes"> Draws on apace: <reg o="foure"> four happy <reg o="daies"> days bring in Another Moon: but oh, me <reg o="thinkes"> thinks, how slow This old Moon wanes; She lingers my desires Like to a Step-dame, or a Dowager, Long withering out a <reg o="yong"> young <reg m="mans"> man's <reg o="reuennew"> revenue.

Hip. <reg o="Foure"> Four <reg o="daies"> days <reg o="wil"> will quickly steep <reg o="thselues"> themselves in nights; <reg o="foure"> Four nights <reg o="wil"> will quickly <reg o="dreame"> dream away the time: And then the <reg o="Moone"> Moon, like to a <reg o="siluer"> silver bow, Now bent in <reg o="heauen"> heaven, <reg o="shal"> shall behold the night Of our solemnities.

Questions?

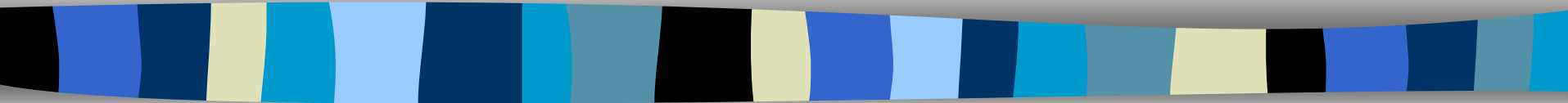


Further information at

<http://www.comp.lancs.ac.uk/ucrel/usas/>

{p.rayson, s.piao. d.archer}@lancs.ac.uk

Appendix



Wmatrix work areas





Wmatrix - Netscape 6


File Edit View Search Go Bookmarks Tasks Help

Dynamic viewpoints: Root: Linguist: Quality: Revere: Summary: Standards: Project: Reader: **Help:** [Introduction](#) [POS tagset](#) [Semantic tagset](#) [Lexicon](#) [Idioms](#)

Wmatrix REVERSE document process: Manual: Workareas: Frequency lists:
Logged in as icame [Tag wizard](#) [Load file](#) [Show all](#) [Show all](#)
[Summary sheet](#) [Create workarea](#) [Show all in detail](#)
[View frequency lists and contexts](#) [LL Wizard](#) [Join](#)

Browse data uploaded

 [TRASH](#)  [LOBM](#)  [sun](#)  [guardian](#)

 [guardian2](#)

- Any items deleted will be moved to the [TRASH](#) area.

If you have technical problems please get in touch with [Paul Rayson](#) (email: paul@comp.lancs.ac.uk)

Document: Done (3.195 secs)

Wmatrix tag wizard

Wmatrix - Netscape 6

File Edit View Search Go Bookmarks Tasks Help

Dynamic viewpoints: Root: Linguist: Quality: Revere: Summary: Standards: Project: Reader: **Help:** [Introduction](#) [POS tagset](#) [Semantic tagset](#) [Lexicon](#) [Idioms](#)

Wmatrix Logged in as icame

REVERE document process: [Tag wizard](#) [Summary sheet](#) [View frequency lists and contexts](#)

Manual: [Load file](#) [Create workarea](#) [LL Wizard](#)

Workareas: [Show all](#) [Show all in detail](#) [Join](#)

Frequency lists: [Show all](#)

Wmatrix tag wizard

Upload file → Part-of-speech tagging → Semantic tagging → Frequency lists

The Wmatrix tag wizard takes you automatically through the POS and Semantic tagging stages, and produces frequency lists from your text file.

Choose file type: Raw text

Choose workarea: New workarea

Or enter new workarea name:

Click the browse button to select a **ascii text file**:

- Raw files will have their contents enclosed in a <text> tag to enable CLAWS tagging.
- If you do not specify a workarea, one will be created with a unique name.









Document: Done (3,224 secs)

Wmatrix views

Wmatrix - Netscape 6

File Edit View Search Go Bookmarks Tasks Help

View of workarea LOBM

File operations	File	Type	Operations
Delete Rename		Raw text	Context for: Word
Delete Rename		SEMTAG output Semantically tagged	Context by: Word POS Semantic Context for: Personal names Modal verbs Proper nouns
Delete Rename		Semantic Frequency list	List: All Compare to normative: BNC IT
Delete Rename		Word Frequency list	List: All Acronyms Section numbers Compare to: <input type="text" value="BNC Sampler Spoken"/> <input type="button" value="Go"/>
Delete Rename		Word-POS Frequency list	List: All
Delete Rename		CLAWS vertical output POS tagged	Run: LEMMINGS CONVERT (to horizontal)
Delete Rename		Word-Sem Frequency list	List: All
Delete Rename		POS Frequency list	List: All Compare to: <input type="text" value="BNC Sampler Spoken"/> <input type="button" value="Go"/>

Document: Done (3,275 secs)

Wmatrix key items

Wmatrix - Netscape 6

File Edit View Search Go Bookmarks Tasks Help

Wmatrix summary sheet for workarea LOBM

Corpus Analysis

The top 7 most significant overused semantic categories are: [\(full list\)](#)

List	Context	Count	Description
List	Context	Z8	1353 Pronouns etc.
List	Context	B1	162 Anatomy and physiology
List	Context	S2.2	46 People:- Male
List	Context	M1	215 Moving, coming and going
List	Context	M6	179 Location and direction
List	Context	X3.2	34 Sensory:- Sound
List	Context	Z1	177 Personal names

Change cut off:

The top 7 words in each of these categories are:

Category	Full List	Top 7 Words
Z8	full list	it he his that you we they
B1	full list	head eyes face arms hand feet membrane
S2.2	full list	Mr men man Mr. boy gentleman fellow
M1	full list	left go passed went returned came come
M6	full list	here away back forward end out beyond
X3.2	full list	sound listening noise whined hear sounded popped
Z1	full list	Steve Heather Dan Harry Edwards Caine Sally

Document: Done (4.166 secs)

Wmatrix frequency lists

View file LOBM.sgm.raw.pos.sem.was.fql

List filtered by string: 'L2_'.
Enter string to limit profile by:

Enter regular expression to limit profile by:

(enter . or nothing for complete list)

Enter name to keep this search
in your user defined list:

pike	L2	12	Context
wings	L2	9	Context
creatures	L2	4	Context
trout	L2	3	Context
shell	L2	3	Context
dog	L2	3	Context
fish	L2	2	Context
bird	L2	2	Context
organism	L2	1	Context
mules	L2	1	Context
lion	L2	1	Context
crow	L2	1	Context
eagle	L2	1	Context
flock	L2	1	Context
jaguar	L2	1	Context
herd	L2	1	Context
cows	L2	1	Context
tail	L2	1	Context
snakes	L2	1	Context
feather	L2	1	Context
hunt	L2	1	Context
webs	L2	1	Context
web	L2	1	Context

Document: Done (0.421 secs)

Wmatrix KIIC

The screenshot shows a Netscape 6 browser window titled "Wmatrix - Netscape 6". The address bar is empty. The menu bar includes "File", "Edit", "View", "Search", "Go", "Bookmarks", "Tasks", and "Help". The main content area displays "KIIC context results". Below this, there is a text input field containing "w1" and a "Save" button. A summary bar indicates "Wrote 44 occurrences." and an "Extend context" button. The main text area shows a search result with a vertical scrollbar on the left. The text is a mix of words and phrases, including "ather stuporous perceptions", "ough all the tuning points of the", "r last time the American felt the", "rue ? Have n't you free speech on", "n , if ever , he would be back on", "t they would be in the next one ,", "eir actions , and their lunar and", "ics which had grown up during the", "ew less humid and as there was no", "umid and as there was no moon the", "ic handiwork . <quote> All those", "rose and returned into the starry", "e snowfields above , not only the", "f course . But perhaps in the new", "n roads , it disappeared from the", "ttle to be seen , only a few lone", "making whoopee under a brilliant", "and read the message aloud . <p>", "midable barter-value on a certain", "world", "world", "Moon", "Moon", "Moon", "Moon", "earthly", "Moon", "moon", "galaxies", "stars", "sky", "sky", "world", "earthly", "stars", "sun", "Terran", "planet", "was an endless cloud in which he f", "'s great radio stations . The dial", "as an outsize social organism whic", "? </quote> <p> <quote> Of course w", "again , and able to do research at", "and her allies being more inferior", "repercussions , to luck (or to Mo", "'s first two centuries ? <p> Harry", "the galaxies had it their own way", "had it their own way so that the f", ". But I 'm a bit earthbound tonigh", "the Indians , if Indians they were", "but the invisible ground of being", "she was entering there might be bo", "scene . <p> The first intimation t", ", and the distant earth brooding i", ". This evening , when First Office", "Headquarters to Bustler . Remain S", "ndash: had not been stolen from s".

At the bottom of the browser window, the status bar shows "Document: Done (4.928 secs)".

A1	GENERAL AND ABSTRACT TERMS	A13.7	Degree: Minimizers	I4	Industry
A1.1.1	General actions, making etc.	A14	Exclusivizers/particularizers	K1	Entertainment generally
A1.1.2	Damaging and destroying	A15	Safety/Danger	K2	Music and related activities
A1.2	Suitability	B1	Anatomy and physiology	K3	Recorded sound etc.
A1.3	Caution	B2	Health and disease	K4	Drama, the theatre and showbusiness
A1.4	Chance, luck	B3	medicines and medical treatment	K5	Sports and games generally
A1.5	Use	B4	Cleaning and personal care	K5.1	Sports
A1.5.1	Using	B5	Clothes and personal belongings	K5.2	Games
A1.5.2	Usefulness	C1	Arts and crafts	K6	Childrens games and toys
A1.6	Physical/mental	E1	EMOTIONAL ACTIONS, STATES AND PROCESSES General	L1	Life and living things
A1.7	Constraint	E2	Liking	L2	Living creatures generally
A1.8	Inclusion/Exclusion	E3	Calm/Violent/Angry	L3	Plants
A1.9	Avoiding	E4	Happy/sad	M1	Moving, coming and going
A2	Affect	E4.1	Happy/sad: Happy	M2	Putting, taking, pulling, pushing, transporting
A2.1	Affect:- Modify, change	E4.2	Happy/sad: Contentment		&c.
A2.2	Affect:- Cause/Connected	E5	Fear/bravery/shock	M3	Vehicles and transport on land
A3	Being	E6	Worry, concern, confident	M4	Shipping, swimming etc.
A4	Classification	F1	Food	M5	Aircraft and flying
A4.1	Generally kinds, groups, examples	F2	Drinks	M6	Location and direction
A4.2	Particular/general; detail	F3	Cigarettes and drugs	M7	Places
A5	Evaluation	F4	Farming & Horticulture	M8	Remaining/stationary
A5.1	Evaluation:- Good/bad	G1	Government, Politics and elections	N1	Numbers
A5.2	Evaluation:- True/false	G1.1	Government etc.	N2	Mathematics
A5.3	Evaluation:- Accuracy	G1.2	Politics	N3	Measurement
A5.4	Evaluation:- Authenticity	G2	Crime, law and order	N3.1	Measurement: General
A6	Comparing	G2.1	Crime, law and order: Law and order	N3.2	Measurement: Size
A6.1	Comparing:- Similar/different	G2.2	General ethics	N3.3	Measurement: Distance
A6.2	Comparing:- Usual/unusual	G3	Warfare, defence and the army; weapons	N3.4	Measurement: Volume
A6.3	Comparing:- Variety	H1	Architecture and kinds of houses and buildings	N3.5	Measurement: Weight
A7	Definite (+ modals)	H2	Parts of buildings	N3.6	Measurement: Area
A8	Seem	H3	Areas around or near houses	N3.7	Measurement: Length & height
A9	Getting and giving; possession	H4	Residence	N3.8	Measurement: Speed
A10	Open/closed; Hiding/Hidden; Finding; Showing	H5	Furniture and household fittings	N4	Linear order
A11	Importance	I1	Money generally	N5	Quantities
A11.1	Importance: Important	I1.1	Money: Affluence	N5.1	Entirity; maximum
A11.2	Importance: Noticeability	I1.2	Money: Debts	N5.2	Exceeding; waste
A12	Easy/difficult	I1.3	Money: Price	N6	Frequency etc.
A13	Degree	I2	Business	O1	Substances and materials generally
A13.1	Degree: Non-specific	I2.1	Business: Generally	O1.1	Substances and materials generally: Solid
A13.2	Degree: Maximizers	I2.2	Business: Selling	O1.2	Substances and materials generally: Liquid
A13.3	Degree: Boosters	I3	Work and employment	O1.3	Substances and materials generally: Gas
A13.4	Degree: Approximators	I3.1	Work and employment: Generally	O2	Objects generally
A13.5	Degree: Compromisers	I3.2	Work and employment: Professionalism	O3	Electricity and electrical equipment
A13.6	Degree: Diminishers			O4	Physical attributes
				O4.1	General appearance and physical properties

O4.2	Judgement of appearance (pretty etc.)	S6	Obligation and necessity	X6	Deciding
O4.3	Colour and colour patterns	S7	Power relationship	X7	Wanting; planning; choosing
O4.4	Shape	S7.1	Power, organizing	X8	Trying
O4.5	Texture	S7.2	Respect	X9	Ability
O4.6	Temperature	S7.3	Competition	X9.1	Ability:- Ability, intelligence
P1	Education in general	S7.4	Permission	X9.2	Ability:- Success and failure
Q1	LINGUISTIC ACTIONS, STATES AND PROCESSES; COMMUNICATION	S8	Helping/hindering	Y1	Science and technology in general
Q1.1	LINGUISTIC ACTIONS, STATES AND PROCESSES; COMMUNICATION	S9	Religion and the supernatural	Y2	Information technology and computing
Q1.2	Paper documents and writing	T1	Time	Z0	Unmatched proper noun
Q1.3	Telecommunications	T1.1	Time: General	Z1	Personal names
Q2	Speech acts	T1.1.1	Time: General: Past	Z2	Geographical names
Q2.1	Speech etc:- Communicative	T1.1.2	Time: General: Present; simultaneous	Z3	Other proper names
Q2.2	Speech acts	T1.1.3	Time: General: Future	Z4	Discourse Bin
Q3	Language, speech and grammar	T1.2	Time: Momentary	Z5	Grammatical bin
Q4	The Media	T1.3	Time: Period	Z6	Negative
Q4.1	The Media:- Books	T2	Time: Beginning and ending	Z7	If
Q4.2	The Media:- Newspapers etc.	T3	Time: Old, new and young; age	Z8	Pronouns etc.
Q4.3	The Media:- TV, Radio and Cinema	T4	Time: Early/late	Z9	Trash can
S1	SOCIAL ACTIONS, STATES AND PROCESSES	W1	The universe	Z99	Unmatched
S1.1	SOCIAL ACTIONS, STATES AND PROCESSES	W2	Light		
S1.1.1	SOCIAL ACTIONS, STATES AND PROCESSES	W3	Geographical terms		
S1.1.2	Reciprocity	W4	Weather		
S1.1.3	Participation	W5	Green issues		
S1.1.4	Deserve etc.	X1	PSYCHOLOGICAL ACTIONS, STATES AND PROCESSES		
S1.2	Personality traits	X2	Mental actions and processes		
S1.2.1	Approachability and Friendliness	X2.1	Thought, belief		
S1.2.2	Avarice	X2.2	Knowledge		
S1.2.3	Egoism	X2.3	Learn		
S1.2.4	Politeness	X2.4	Investigate, examine, test, search		
S1.2.5	Toughness; strong/weak	X2.5	Understand		
S1.2.6	Sensible	X2.6	Expect		
S2	People	X3	Sensory		
S2.1	People:- Female	X3.1	Sensory:- Taste		
S2.2	People:- Male	X3.2	Sensory:- Sound		
S3	Relationship	X3.3	Sensory:- Touch		
S3.1	Relationship: General	X3.4	Sensory:- Sight		
S3.2	Relationship: Intimate/sexual	X3.5	Sensory:- Smell		
S4	Kin	X4	Mental object		
S5	Groups and affiliation	X4.1	Mental object:- Conceptual object		
		X4.2	Mental object:- Means, method		
		X5	Attention		
		X5.1	Attention		
		X5.2	Interest/boredom/excited/energetic		