A Medical Note

Automated Parsing
And
Semantic Extraction

The Note

Sections show a moderately differentiated squamous cell carcinoma with cystic change. The tumour extends to the visceral pleura but unequivocal pleural invasion is not seen in the routinely stained sections. An elastic stain has been requested for further evaluation of the pleura and a supplementary report will be issued with the result. Lymphatic invasion is not identified and blood vessel invasion is not seen. The bronchial resection margin shows squamous metaplasia but no evidence of in situ or invasive carcinoma. The vascular resection margin is negative for carcinoma. Non-neoplastic lung tissue shows centrilobular emphysema and smoker's-type macrophages. A section from the tip of the lobe confirms the presence of honeycomb change and shows associated bronchiolar metaplasia. Anthracotic dust macules are seen with associated emphysema. Additional sections of non-neoplastic lung tissue will be examined for further evaluation.

Microscopy shows profiles of lymph node tissue. Metastatic carcinomas is not identified.

From AEHRC website

Degree of Difficulty

The sentences are short and crisp, making parsing pretty easy in comparison with sentences in specifications or contracts.

That doesn't mean that complex concepts aren't being linked together, but it does mean the complexity is in the concepts, rather than the linking, and no new concept is being introduced.

The objects are mostly concrete, at least while we stick to physical medicine.
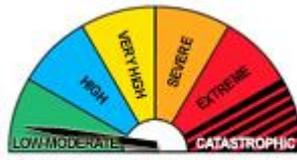
Integrated Semantics

Semantics is integrated into the grammatical patterns – you don't get very far without knowing something about the objects being referred to in the text.

A word can be a verb or a participle or an adjective or a noun and the neighbouring words may not be in context, so the system needs to know what the word means to figure out its purpose in the context of the sentence.

Frequent references are made to the meaning of words while parsing is proceeding.
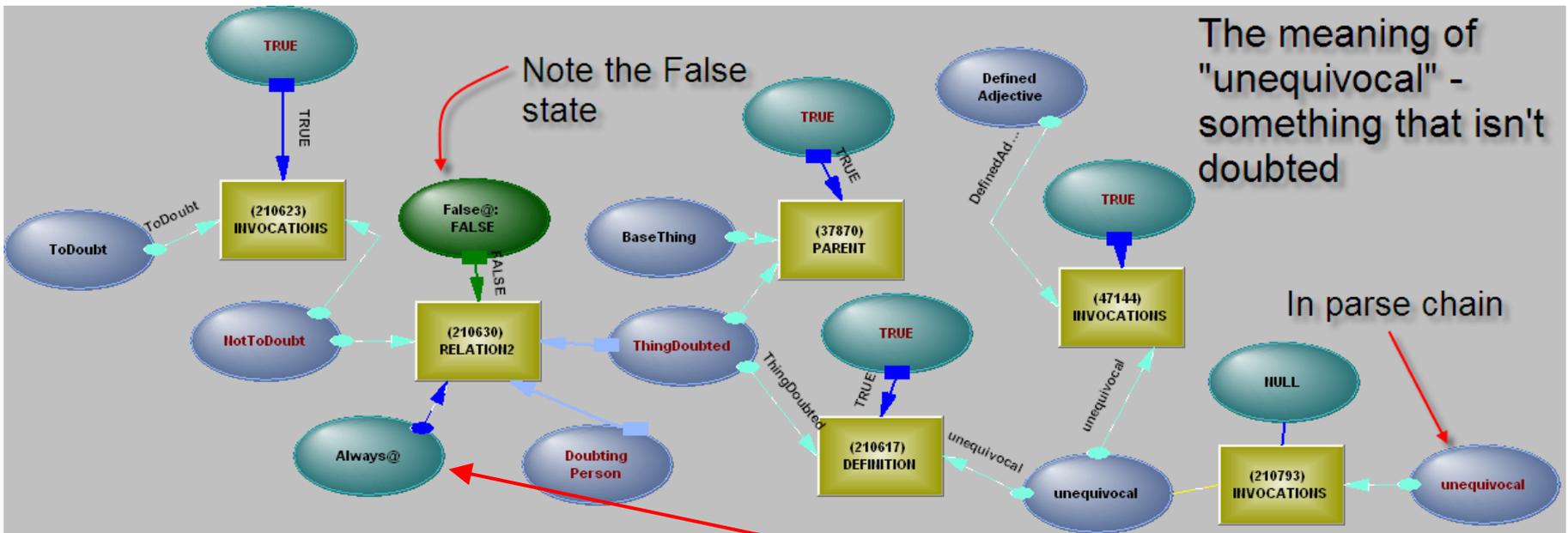
"indubitable" would have a false on the existential pin – impossible to doubt

Accurate Definitions

Complete and accurate definitions of medical terms would seem to be essential if the system is to be useful in semantically analysing medical notes – otherwise all it is doing is connecting words it knows nothing about.

Some terms describe a process – either a living process or a surgical procedure, with multiple actions and states and slippery phrases like "different" or "of another type".

The following two slides show how concepts are turned into detailed descriptions, using a medical dictionary as a reference and a combination of relations and states.

Definition of Metastasis

Main Entry: **me·tas·ta·sis** 🔊

Pronunciation: \mə-'tas-tə-səs\

Function: *noun*

Inflected Form(s): *plural* **me·tas·ta·ses** \-,sēz\ 🔊

1 a : change of position, state, or form b : the spread of a disease-producing agency
(as cancer cells or bacteria) from the initial or primary site of disease to another part
of the body <*metastases* of breast cancer to bone—*Medical Physics*> ; *also* : the
process by which such spreading occurs

2 : a secondary malignant tumor resulting from metastasis

The structural definition is limited to only cancer.

There are two meanings – the process of spreading or the resultant tumor.

(the FROMTO is used for all movement relations)

Metastasis & Metastatic

Definition of Metaplasia

Main Entry: **meta·pla·sia** 🔊

Pronunciation: \,met-ə-'plā-zh(ē-)ə\

Function: *noun*

1 : transformation of one tissue into another <*metaplasia* of cartilage into bone>

2 : abnormal replacement of cells of one type by cells of another

Meaning 2 isn't very good – it doesn't mean you mistakenly tip some cells out of a test tube and put in different ones. We will stick to transformation, although the meaning probably refers to self-replacement (a good example of why automatic parsing of dictionary entries would be fraught). The use will normally be describing the tissue, rather than the process, so we will give it two meanings

What About the Buts

The tumour extends to the visceral pleura but unequivocal pleural invasion is not seen in the routinely stained sections.

"but" represents a denial of expectation or other probability –

John thought Fred was guilty, but he was innocent

The "thought" provides a logical expectation, which is rebutted without error.

We looked for evidence
       (BECAUSE we expected to find evidence)
   BUT we did not find evidence

BUT is an operator that handles the clash of logical states without producing an error.

Everything has to be represented faithfully if the semantic structure is to be valid (which may cause you to think hard in giving it form, because you know what it means, you just never had to think at this level of detail).

Parsing

Building of the semantic structure occurs in parallel with parsing so the system can know as soon as possible what the words are actually standing for and prune away irrelevant meanings (and parts of speech).

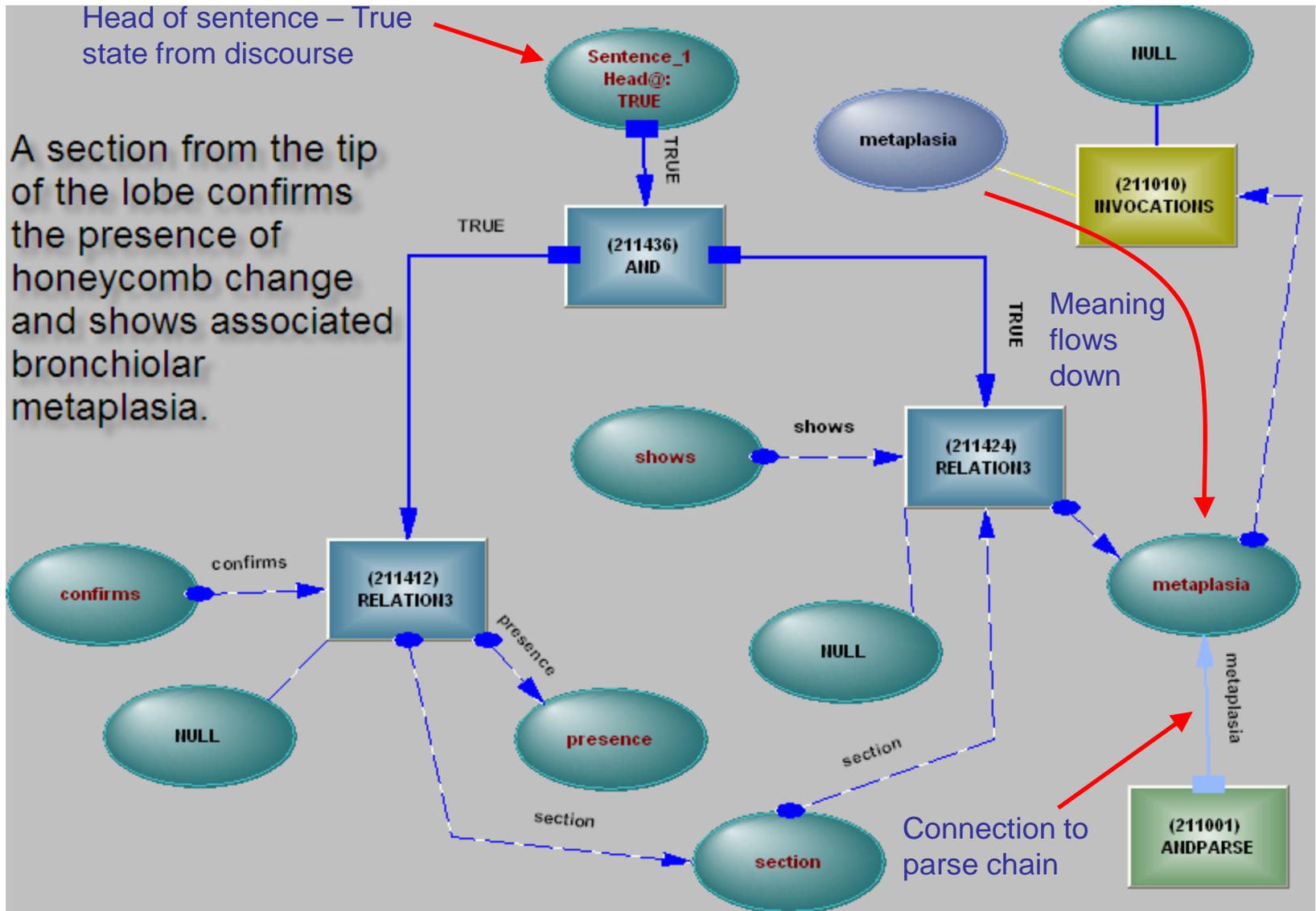We are not sure whether it is a noun or a verb yet so we have to build more to be sure
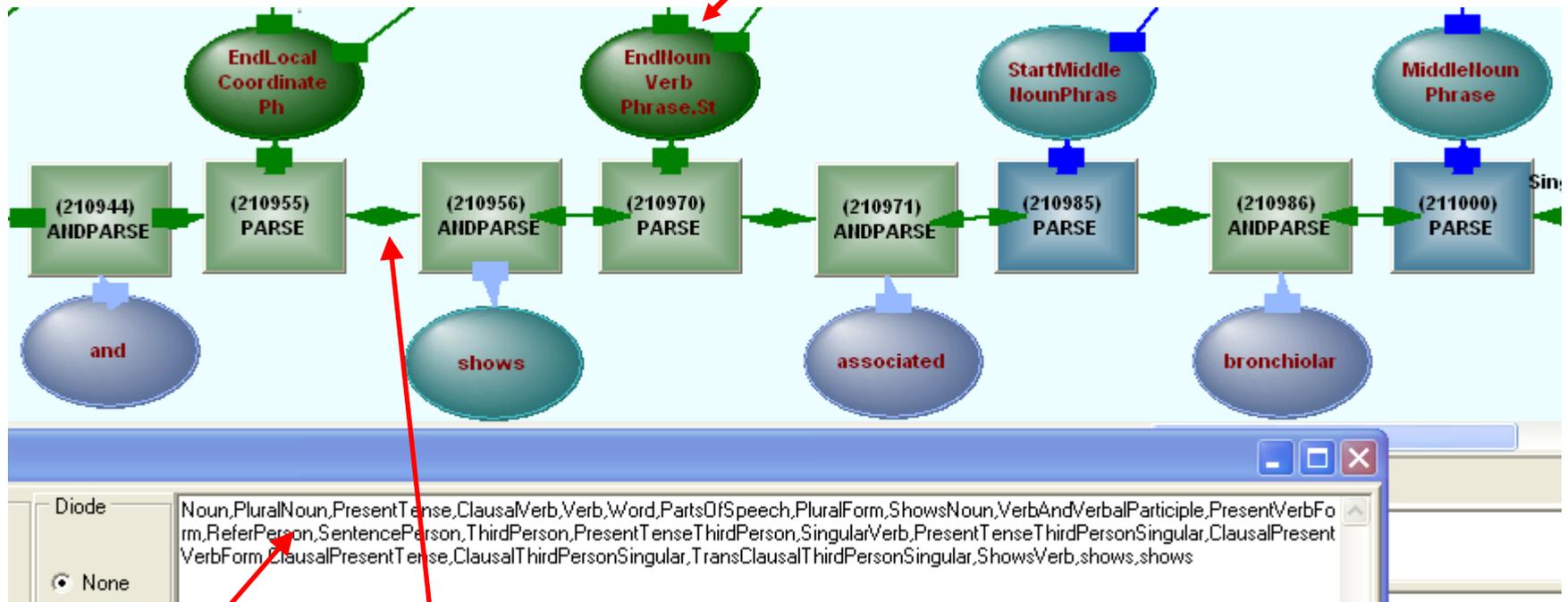
This set is in that link so fast pruning is important – here we couldn't prune it until later

Performance

	Sentence	Imported	Consistent	Time	Discourse	Full Build	Elements
1	Sections show a moderately differentiated squamous cell carcinoma with cystic ch...	0		6	Yes		458
2	The tumor extends to the visceral pleura but unequivocal pleural invasion is not...	0		8	Yes		682
3	An elastic stain has been requested for further evaluation of the pleura and a s...	0		15	Yes		771
4	Lymphatic invasion is not identified and blood vessel invasion is not seen...	0		4	Yes		358
5	The bronchial resection margin shows squamous metaplasia but no evidence of in s...	0		9	Yes		508
6	The vascular resection margin is negative for carcinoma...	0		3	Yes		327
7	Non - neoplastic lung tissue shows centrilobular emphysema and smoker's - type m...	0		5	Yes		379
8	A section from the tip of the lobe confirms the presence of honeycomb change and...	0		17	Yes		719
9	Anthracotic dust macules are seen with associated emphysema...	0		3	Yes		317
10	Additional sections of non - neoplastic lung tissue will be examined for further...	0		5	Yes		485
11	Microscopy shows profiles of lymph gland tissue	0		5	Yes		296
12	Metastatic carcinomas is not identified	0		1	Yes		199
	Total 12			Total 81	Total 12		

Discourse "yes" means the structure was built to the level of the discourse, and connected to it
The network currently is at 510,000 elements and the current limit is 1,600,000 elements
A complete oncology specialty might take it to 1,000,000 elements after salvaging all
grammatical artifacts – averaging about 12 links per word or concept

Semantic Structure 1

Head of sentence – True state from discourse

A section from the tip of the lobe confirms the presence of honeycomb change and shows associated bronchiolar metaplasia.

Sentence_1 Head@: TRUE

NULL

metaplasia

(211010) INVOCATIONS

TRUE

(211436) AND

TRUE

Meaning flows down

TRUE

shows

(211424) RELATION3

confirms

(211412) RELATION3

presence

NULL

metaplasia

NULL

section

Connection to parse chain

metaplasia

(211001) ANDPARSE

section

section

presence

Why Do This?

The resulting semantic structure is searchable using structural search methods, and as it represents all the meaning of the text, more reliance can be placed on the results.

Simple synonyms, rotational and complex synonyms, state conversions, dimensional conversions, generation of sets of possibilities and pruning, all can be used in searching the structure.
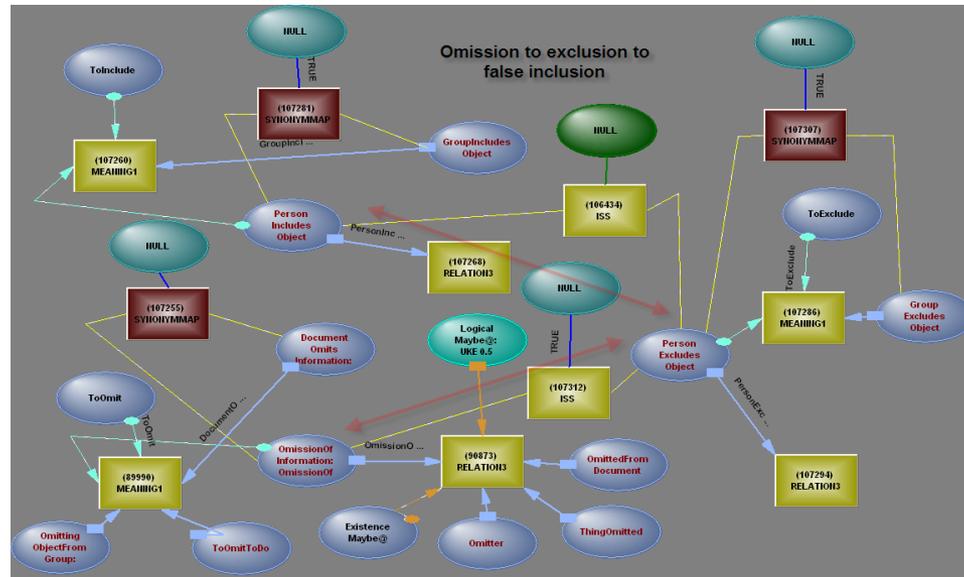
Utility

By automating the collation, analysis, summarisation and classification of relevant patient data (pathologist reports and metadata), the reliance on expert clinical staff can be lessened, improving the efficiency and availability of cancer information.
From AEHRC site

Automatic operations on patient data can be made more accurate and more extensive by using high reliability semantic structure extraction. This would allow the current project goals to be met, and other higher goals to become feasible, including semantic searching for systemic factors that are not predefined.

(if you have limited the semantic bits to be handled, you can't expect to find unexpected things of interest outside those – by making no such presupposition and building all that the note says, and combining the note structure with other structures, many more things can be found, including things that exist in no single structure or require dynamic search construction – structure built during the search in response to what is found along the way).