Pragmatic Annotation

with reference to the Engineering Lecture Corpus (ELC)

Hilary Nesi, Ummul Ahmad & Noor Mala Ibrahim

www.coventry.ac.uk/elc

Lessons from BASE
www.coventry.ac.uk/base

- 160 lectures
- 40 seminars

- Video/audio recordings not yet time-aligned
- Not pragmatically annotated, but studies of the lecture data have revealed major (British) lecture functions

The Engineering Lecture Corpus

Collections of video-recorded lectures on similar topics from

- The UK

- Malaysia

- New Zealand

- And 5 other countries (pending funding) – Spain, Hong Kong, Taiwan, Germany, UAE

Why engineering?

International English-medium education……

Project aims

- To identify and describe typical engineering lecture discourse features

- To compare English-medium engineering lecturing styles in different parts of the world

Why annotation?

- lexical items are easy to find in a corpus. So are grammatical patterns – with POS tagging and an appropriate interface. Starting with a functional category is much more difficult.
- Engineering lectures (and academic lectures generally) share some basic functions, but these may be realised in many different ways.
- Variation in the realisation and distribution of functions is one of the most interesting aspects of the ELC.

Sinclair's views

'Any information about a text other than the alphanumeric string of its words and punctuation should be stored separately from the plain text and merged when required in applications.'

Corpus and Text: Basic Principles (2005)

http://ahds.ac.uk/linguistic-corpora

Pragmatic annotation in MICASE

Decision to mark up pragmatic features that were:

1. not easily searchable by other means
2. prevalent but not ubiquitous
3. Unambiguous

(Maynard & Leicher, 2007)

MICASE Pragmatic Tags

1. ADV   Advice (giving, soliciting)
2. AHW   Assigning homework
3. DEF   Defining/glossing terms
4. DIR   Directives
5. DIS   Disagreement
6. EVN   Evaluation (positive, negative)
7. IRM   Introductory roadmap
8. NAR   Narrative
9. Q   Questions
10. REQ   Requests
11. SPI   Speaker introductions
12. TAN   Tangents

Broad brush approach

- To account for functions in a range of academic speech events

- Some tags refer to functions that are not prevalent in our lecture data (Directives, Disagreement)

- Some functions that are prevalent in our lecture data are not listed.

Rules for ELC annotation

Chosen features should:

- shed light on the specific nature of lecture discourse

- not be recoverable from TEI contextual markup (e.g. <vocal desc ="laughter"> <kinesic desc="writes on board">)

- not realised by a predictable form (e.g thanks / thank you – the frequency of which is nevertheless quite interesting to compare across national groups)

- be likely to occur more than once.

ELC Pragmatic Tags (provisional)

1. Greetings
2. Prayer
3. Housekeeping
4. Preview / Review
5. Advice
6. Defining
7. Reference to ss future profession
8. Personal narrative
9. Teasing
10. Self-recovery
11. Self-deprecation
12. Black humour
13. Disparagement of out-group member
14. Register / word play

MICASE v ELC

ADV Advice (giving, soliciting)

AHW Assigning homework

DEF Defining/glossing terms

DIR Directives

DIS Disagreement

EVN Evaluation (positive, negative)

IRM Introductory roadmap

NAR Narrative

Q Questions

REQ Requests

SPI Speaker introductions

TAN Tangents

- ✓
- ✓ 'Housekeeping' is broader
- ✓ May involve code-switching
- N/A
- N/A
- ? Possible candidate

- ✓'Preview' (but also 'Review')
- ✓ Personal narratives
- N/A – fairly easy to identify
- N/A
- N/A
- N/A

Additional ELC tags

- Greetings
- Prayer
- Review (content of current or past lecture)
- Reference to future profession
- Teasing
- Self-recovery
- Self-deprecation
- Black humour
- Disparagement of out-group member
- Register/word play

Some examples

- Similarities and differences across the British, Malaysian and New Zealand data

Greetings

- assalamualaikum and good afternoon

    (MCiv001)

- right folks it's ten past time to start morning nice and cold today lovely lovely bit of a nip in the air um right (CCiv005)

But no greetings (yet) in NZ data -

- Just before we get started we're going to look today at working in groups ……..(NZDMech002)

Some teasing

- right everyone happy or asleep Monday morning wakey wakey (NZSolidMec001)
- if it is too comfortable then perhaps half of the class will be falling asleep (MMec001)
- i'm just going to work my way clockwise around the frame (RINGTONE, LAUGHTER) sorry is im just going to do a bit of patricide (CCiv005)

Defining – UK data

- bleeding is when water comes out of the mix it's a particular part of **what we call segregation** when <u>we make</u> concrete <u>we mix</u> many ingredients together as I've said the cement the aggregates the water um **segregation is when they all come apart** when <u>we've placed them</u> cos they've all got different densities

<div align="right">CCiv010</div>

- if we use this thing here erm in this region here you see that the it's a funny sort of curve <u>you got here</u> erm in this region is where <u>your neutral axis</u> is erm within the section (pause) and **your tension reinforcement or the reinforcement at the bottom of the section if you like** on the bottom face as <u>we're looking at it</u> erm that yields and then <u>you get</u> crushing of the concrete on the top section **so this is what's called erm tension failure**

<div align="right">CCiv017</div>

Defining- NZ data

- now voltage is between two points so I can have **lines here which we actually call equapotentials**

<div align="right">(NZElec004)</div>

- <u>what you will find</u> is that <u>you take</u> a tensile specimen you put it into a tensometer which <u>you are going to do</u> some labs in across the way <u>you put it you measure</u> the force <u>you measure</u> the area and eventually at some point it breaks **that breaking point and the stress of the breaking point is known as the ultimate tensile stress**

<div align="right">(NZMec001)</div>

Defining – Malaysian data

G K actually is meant for uh dead load dead load coming from dead load coming from the slab slab *lah **beban mati** eh tak ada bergerak* it's slab uh the Q K uh the G K coming from the slab and then the Q K is is er **life load sometimes we call it epos load** okay the life load or the epos load all given in the in the data here actually all given   (MCiv002)

so by the definition **moment of a force above a point or an axis is the measure of the tendency of a force to force a body to rotate about the point of the axis**
                                                                    (MCiv001)

Self deprecation

- those of you who were lucky enough to have hair when you comb your hair in the morning on a dry day what you find is that sometimes the hair sticks to the comb that happens have you ever had that I've got fond memories of that happening <laughter> (NZEIec001)

Black humour in M ELC?

- in Japan there is a term called karoshi this is where the employees just collapse and die at the work place there's even a term for that call karoshi but we don't have that term in our language because it's it's it does not probably seldom happen or probably people people uh still the work stress here is still not as high as Japan but yah in Japan they call it karoshi that mean death attributed to uh stress at the work place so just like me come here and teach and collapse and pass away (MMec001)

Black humour in UK ELC?

- What will happen then is I will open it up again for another two weeks except for the person who's phone's going off <laughter> cos they're not gonna be able to sit down for about a month <laughter> yes I am hard…. Thank you very much are we done switched off or inserted on vibrate <laughter> (CCiv004)

Prospects

- Current funding from a British Council PMI2 Research Cooperation grant, the Malaysian government, and Auckland University of Technology
- Funding applied for from the Leverhulme Trust
- All future partners have already received local funding

References

- Sinclair, J. (2005) Corpus and Text: Basic Principles. In Wynne, M. (ed) *Developing Linguistic Corpora: a Guide to Good Practice* http://ahds.ac.uk/linguistic-corpora

- Jurafsky, D. Pragmatics and computational linguistics. In Horn, L. & Ward, G. (eds) *Handbook of Pragmatics*. Oxford: Blackwell

- Maynard, C. & Leicher, S. (2007) Pragmatic annotation of an academic spoken corpus for pedagogical purposes. In Fitzpatrick, E. (ed) *Corpus Linguistics Beyond the Word: Corpus Research from Phrase to Discourse.* Amsterdam: Rodopi

- Swales, J. (2005) How to do MICASE-based investigations. http://lw.lsa.umich.edu/eli/micase/How_to_function.htm