Developing an Annotation Scheme

- Motivation
- Basic scheme
- Preliminary Annotation
- Informal evaluation & development
- Scheme Revision and re-coding
- Coding manual
- Formal evaluation: inter-coder reliability

⇨ Ready to code real data

Motivation:

- **Question**
  - Might be very specific, or more general
- **E.g. What kind of dialogue acts are there**
  - (in a particular genre of dialogue)
  - (that perform a particular type of function)

Basic Scheme

- **Preliminary categories that seem to cover the range of phenomena of interest**
  - Different categories functionally important and/or easy to distinguish

Dialogue Act Taxonomy considerations

- **How detailed?**
  - difference in conditions/effects vs. confidence in label
  - capture generalizations or distinctions?
    - example: state, assert, inform, confess, concede, maintain, affirm, claim,...
- **Where should complexity reside?**
  - Multi-functional, complex acts?
    - Possibly many acts
    - Possibly performances that can not be labelled
    - Ex: verbmobil 1
  - Many (simple) acts per performance
    - Possibly many tagging decisions
    - Ex: Damsl/DRI

corpus annotation comparisons

- **Activities**
  - Trains movement planning (Trains)
  - disaster relief planning (Monroe)
  - Casual conversation (Switchboard)
  - Maptask
  - Scheduling appointments (Verbmobil)
- **Participants**
  - Language (English vs German)
  - Organizational status (students (HCRC) vs military (DCIEM)
- **Dialogue act taxonomies**
  - HCRC
  - Verbmobil (I & II)
  - Damsl
  - SWBD-Damsl

Distribution of dialogue acts in corpora

Damsl TRAINS	Damsl Monroe	SWBD-Damsl Switchboard	HCRC HCRC Maptask	HCRC DCIEM Maptask	Verbmobil II Verbmobil English	Verbmobil II Verbmobil German	Verbmobil I Verbmobil I German
statement 45.9	51.4	49	explain 7.9	7.9	Inform,... 22.8	21.2	12.2
info-request 15.2	9.9	questions 4.9	query,check,align 23.5	20.3			
action-dir,oo 12.2	12.9	0.7	instruct 15.6	15.2	request,suggest 26.0	27.0	32
commit,offer 23.8	16.8	0.1			commit 0.5	0.8	
conventional 2.5	0.6	1.4			13.4	15.6	16.5
answer 14.7	8.4	3	reply,clarify 22.8	20	feedback 15.2	9.8	0.6
accept 30.0	23.0	5			accept,confirm 10.3	12.3	13.5
reject 2.2	0.5	0.2			reject,explained 3.3	4.4	8.2
other agree 3.6	1.8	0.3			clarify 2.3	1.9	8.9
Understanding 30.2	28.5	23	acknowledge 20.5	28.1	backchannel 3.6	3.3	
non-understand 1.2	0.5	0.1					

Dialogue Diversity

- **LDC**
- **Allwood: The Swedish Spoken Language Corpus at Goteborg: multiple activities**
  - http://www.ling.gu.se/projekt/tal/
- **Mann: Dialogue diversity corpus**

  http://www-rcf.usc.edu/~billmann/diversity/DDivers-site.htm

Taxonomy principles:

- **Activity-specific**
  - Must cover activity features
  - Make crucial distinctions
  - Avoid irrelevant distinctions (reduce perplexity)
- **General**
  - Aim to cover all activities
  - Specific activities work in a sub-space
  - Activity-specific clusters as "macros"

Types of Dialogue

- **Task-oriented:**
  - dialogue about a task performance
- **Information-oriented:**
  - one participant needs information that others have
- **Relationship-oriented:**
  - purpose is influence the nature of the relationship (become closer, establish trust, expertise or dominance)
- **Individual-oriented:**
  - (someone "wants to talk", express self, listener effects not important)

Preliminary Annotation

- **Algorithm**
  - Automated annotation if possible
    - Semi-automated
      - Partial
      - Supervised decisions
  - Decision trees for human annotators
- **Definitions, guidelines**
- **Multiple annotators**
  - Ideally following official guidelines or algorithm rather than informally taught

Informal evaluation & development

- **Analysis of problematic annotations**
  - Are some categories missing?
  - Are some categories indistinguishable for some coding decisions?
  - Do categories overlap (is this allowed)
- **Meetings between annotators and scheme designers and users**
- **Revision of annotation guidelines**
- **More annotation**

⇨ **Annotation manual**

Formal evaluation

- **Controlled coding procedures**
  - Individuals coding unseen data
  - Coding on the basis of manual
  - No discussion between coders
- **Evaluation of inter-coder reliability**
  - Confusion matrix
  - Overall