



Normalizing dialogue act tagsets: how to keep ISO standards general

Andrei Popescu-Belis

ISSCO, University of Geneva

Sixth Workshop on Multimodal Semantic Representation
ACL-SIGSEM WG on the Representation of Multimodal Semantic Information
& ISO TC37/SC4 TDG3 on Semantic Content Representation.
8-9 January 2007, Tilburg, The Netherlands

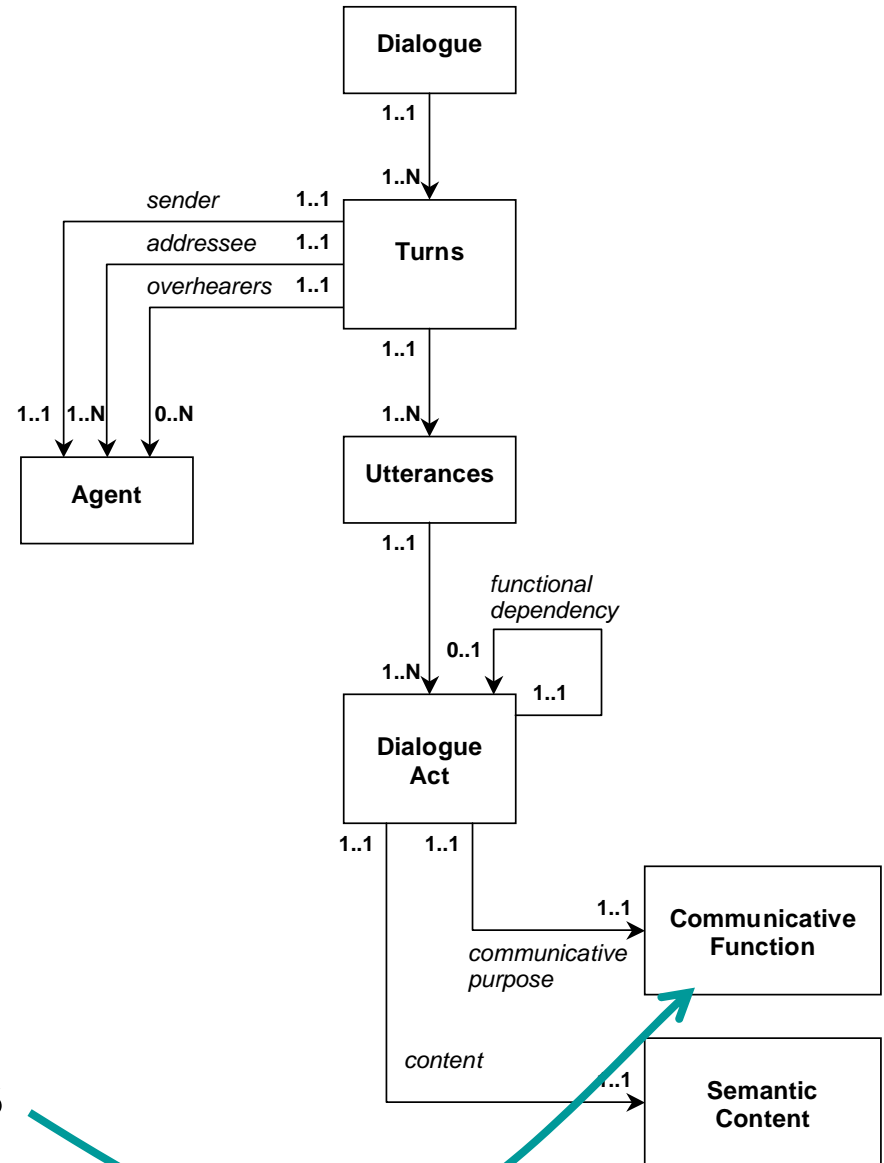
Observations from previous research on dialogue acts (DAs)

- Many theoretical frameworks
 - linguistics/pragmatics
 - computational linguistics
 - dialogue systems
 - Many DA tagsets
 - see for instance the MATE report on DAs
 - **Commonalities?**
 - idea that utterances have semantic content and **communicative functions** → attempts to **taxonomize** CFs
 - But a lot of **differences** between these attempts...
-

Segmentation

SOURCE: “Preliminary metamodel for dialogue acts” (H. Bunt, 2006 ACL-SIGSEM talks)

- There seems to be a wide agreement on such a model
 - maybe some issues in case of overlaps
- This talk focuses on communicative functions



Plan

1. Origins of my interest in DAs
 2. Theoretical inspirations
 3. Overview of some DA tagsets
 4. What could an ISO standard on DAs contain?
 - tagsets, or principles for defining tagsets?
-

My interest in DAs (1/2)

■ Institutional aspects

- (IM)2, Swiss project on “Interactive Multimodal Information Management”, 2002-2009
- AMI, “Augmented Multiparty Interaction”, EU FP6 IP, 2004-2006 (follow-up: AMIDA, 2006-2009)
 - same leading house: IDIAP, Martigny
- ISSCO: (IM)2 partner working on Multimodal Dialogue Management & Content Abstraction

■ Research issues

- recognize DAs in multiparty dialogues
 - training/test: convert existing dialogues + annotate new ones
 - need to find or to define a convenient DA tagset
-

My interest in DAs (2/2)

- Analysis of tagset “dimensionality”
 - theoretical grounding & search space of automatic taggers
 - *Dialogue act tagsets for meeting understanding: an abstraction based on the DAMSL, Switchboard and ICSI-MR tagsets* IM2 Project Report MDM-09, September 2003, 50 p.
<http://www.issco.unige.ch/projects/im2/mdm/docs/apb-mdm-report09-v12.pdf>
- Definition of the MALTUS tagset
 - smaller search space & compatibility with previous resources
 - “Abstracting a Dialog Act Tagset for Meeting Processing”, *Proc. of LREC 2004*, Lisbon, p.1415-1418
- DA tagging with dimension-specific classifiers
 - pilot experiments with HMM models
 - “Multi-level Dialogue Act Tags”, *Proc. of the 5th SIGdial Workshop*, 2004, Cambridge, MA, p.163-170, (with A. Clark)

What goals do people have when designing DA tagsets?

1. Formalize theoretical assumptions about utterance-functions
 2. Annotate dialogues by hand, for empirical research on DAs
 3. Design automatic DA taggers and measure their performance
 4. Build NLP applications based on DA recognizers / generators
 5. Convert existing resources to new DA tagsets
-

Theoretical inspiration

- There is a significant body of research on taxonomies of utterance-functions (communicative functions) in linguistics, discourse studies, and pragmatics
 - review pages 4-11 of *Dialogue Acts: One or More Dimensions?* ISSCO Working Paper nr. 62, Nov. 2005, 30p.
<http://www.issco.unige.ch/publications/working-papers/>

 - Questions
 - are theories complementary or conflicting?
 - (how) should they be included in an ISO standard?
 - how do actual DA tagsets relate to theories? should they?
-

Some theories of utterance function (1 / 3)

1. **Speech acts** (Austin, Searle, and others)

- assert, suggest, promise, question, ...
- not *the* but *a* function of an utterance

2. **Turn management** (conversation analysis, Schegloff)

- turn-giving, turn-holding, etc.

3. **Adjacency pairs**

- links between a first-part and a second-part utterance such as question and answer
- more than speech acts labels (“answer” is more than a speech act)

Some theories of utterance function (2/3)

4. **Overall organization and topics** (conversation analysis)
 - ❑ conventional openings and closings, series of topics
 - ❑ utterance can serve as an opening or closing, or serve to introduce or to close a topic

 5. **Politeness management** (e.g., Brown and Levinson 1984)
 - ❑ politeness marks appear in almost every utterance
 - ❑ face-threatening / face-preserving role
 - ❑ or simply mark an utterance “politeness”

 6. **Rhetorical role** (e.g., Rhetorical Structure Theory, Mann and Thompson)
 - ❑ assign to each utterance a rhetorical role
 - ❑ less often applied to dialogue (better on monologue?)
-

Some theories of utterance function (3/3)

7. **Dynamic Interpretation Theory (H. Bunt)**

- task-oriented vs. dialogue control functions
- DIT++ tagset

8. **Other integrative theories**

- Traum & Hinkelman 1992 multi-level approach
- Allwood 2000 activity-based approach
- H. Clark 1996: importance of grounding in dialogue

9. **Global theories of communication**

- Discourse Representation Theory
 - Relevance Theory
-

Some DA tagsets

- DAMSL
 - SWBD-DAMSL
 - ICSI-MRDA
 - MALTUS
 - AMI
 - DIT++
-

DAMSL (Allen & Core 1997)

- Dialog Act Markup in Several Layers
 - Multidimensional tag set
 - Information Level (4 non-exclusive tags)
 - Communicative Status (3 n-e tags)
 - Forward Looking Function (7 dimensions with 3-4 n-e tags each)
 - e.g., statement, info-request, committing-speaker-future-action
 - Backward Looking Function (3 dimensions)
 - agreement (6 n-e tags), understanding (6 n-e tags), answer (y/n)
 - Nearly independent tags
 - annotation instruction: “use as many as needed”
 - ~ 4 million combinations
-

Our formalization of DAMSL (simplified) using rewriting rules

DA → **CS?** ^ **IL?** ^ **FLF?** ^ **BLF?**

CS → uninterpretable? ^ abandoned? ^ self-talk?

IL → task | task-management | communication-management | other-level

FLF → statement? ^ influencing-addressee-future-action? ^ info-request? ^
committing-speaker-future-action? ^ conventional? ^
explicit-performative? ^ exclamation? ^ other-ff?

statement → assert | reassert | other

influencing-addressee-future-action → open-option | action-directive

committing-speaker-future-action → offer | commit

conventional → opening | closing

BLF → agreement? ^ understanding? ^ answer?

agreement → accept | accept-part | maybe | reject-part | reject | hold

understanding → signal-non-understanding | acknowledge |

repeat-rephrase | completion | correct-misspeaking

SWBD-DAMSL (Jurafsky et al. 1998)

- “Application” of DAMSL to the Switchboard data (Jurafsky et al. 1997)
 - telephone conversations >200'000 utterances
 - reduce dimensionality by observing which DAMSL tags never appear together
 - or: keep most frequent combinations
- Result: 220 unique tags → clustered into 42 mutually-exclusive tags
- Examples of most frequent tags
 - Statement 36%
 - Continuer 19%
 - Opinion 13%
 - Agree/Accept 5%
 - Abandoned/Turn-Exit 5%
 - Appreciation 2%
 - Yes-No-Question 2%
 - Non-verbal 2%
 - Yes answers 1%
 - Conventional-closing 1%
 - Uninterpretable 1%
 - Wh-Question 1%
 - No answers 1%
 - Response Acknowledgment 1%
 - Hedge 1%
 - Declarative Question 1%
 - Other 1%
 - Backchannel-Question 1%

HB (previous talk): do not forget that these are semantically complex!

ICSI-MRDA (Shriberg et al. 2004)

- ICSI-MR idea: reuse the SWBD-DAMSL tags, but...
 - combine several tags: again multidimensional
 - alter the meaning of some tags
 - some new tags, for multiparty dialogue: e.g., “floor grabber”
 - Definition: ICSI-MR label composed of
 - one **general tag**
 - zero, one or (many) more **specific tags**
 - five **disruption marks**
 - about 60 tags are defined
 - Problem: tag definition (=meaning) vs. ICSI guidelines vs. observed use in the ICSI-MR annotated corpus
-

Our formalization of ICSI-MRDA

$g\text{-tag} [\text{^s-tag1} \dots \text{^s-tagN}].\text{dis-tag}$

$DA \rightarrow DA' \text{^ \%? ^ x?}$

$DA' \rightarrow T1 (\text{^ T2})^* (.T3)?$

$T1 \rightarrow s | Q | b | FM$

$Q \rightarrow qh | qo | qr | qrr | qw | qy$

$FM \rightarrow fg | fh | h$

$T3 \rightarrow \% - | \% --$

$T2 \rightarrow RE? \text{^BA? ^CK* ^RI? ^SF? ^PO? ^FD*}$

$RE \rightarrow RP | RN | RU$

$BA \rightarrow bh | bk$

$AC \rightarrow cc | co | cs | fx$

$CK \rightarrow br | bu | f | g$

$RI \rightarrow RIC | RIR$

$RIC \rightarrow bc | bsc$

$RIR \rightarrow r | bs | m$

$SF \rightarrow ba | df | fe | 2 | e$

$PO \rightarrow bd | by | fa | ft | fw$

$FD \rightarrow rt | t | t3 | tc | t1 | d | j$

MALTUS

- Multidimensional Abstract Layered Tagset for UtteranceS
 - Formation rules for a MALTUS label
 - $DA \rightarrow (U \mid T1 (T2)?) (D)?$
 - $T1 \rightarrow S \mid Q \mid B \mid H$
 - $T2 \rightarrow (RP \mid RN \mid RU)? AT? DO? (RIC \mid RIR)? PO?$
 - Number of theoretically possible combinations: 770
 - Conversion of ICSI-MR tags to MALTUS
-

Definitions of MALTUS labels

- U = undecipherable (unclear, noisy)
- S = statement
- Q = question
- B = backchannel
- H = hold (floor holder, floor grabber, hold)
- RP = positive answer (or positive response)
- RN = negative answer (or negative response)
- RU = other answer (or undecided answer or response)
- RIC = restated information with correction
- RIR = restated information with repetition
- DO = command or other performative (can be refined into: command, commitment, suggestion, open-option, explicit performative)
- AT = the utterance is related to attention management (can be refined into: acknowledgement, rhetorical question backchannel, understanding check, follow me, tag question)
- PO = the utterance is related to politeness (can be refined into sympathy, apology, downplayer, “thanks”, “you're welcome”)
- D = the utterance has been interrupted or abandoned

Number of possible vs. occurring DA labels in the ICSI-MR corpus: two tagsets

Tagset	Tags per label (N)	Theoretical (possible) label types	Actual (observed) label types	Actual (observed) label tokens	Maximum accuracy with N tags
ICSI-MRDA	1	11	11	68,213	0.6007
	2	429	129	37,889	0.9343
	3	8,151	402	5,054	0.9788
	4	100,529	176	2,064	0.9970
	5	904,761	49	326	0.9999
	6	6,333,327	9	14	1.0000
Total	–	7,347,208	776	113,560	–
MALTUS	1	4	4	84,092	0.74051
	2	28	14	28,366	0.99003
	3	72	29	1,089	0.99997
	4	88	3	3	1.00000
Total	–	192	50	113,560	–

Frequent MALTUS labels

- 22 labels occurring more than 20 times each
 - ordered alphabetically
 - for about 110,000 utterances

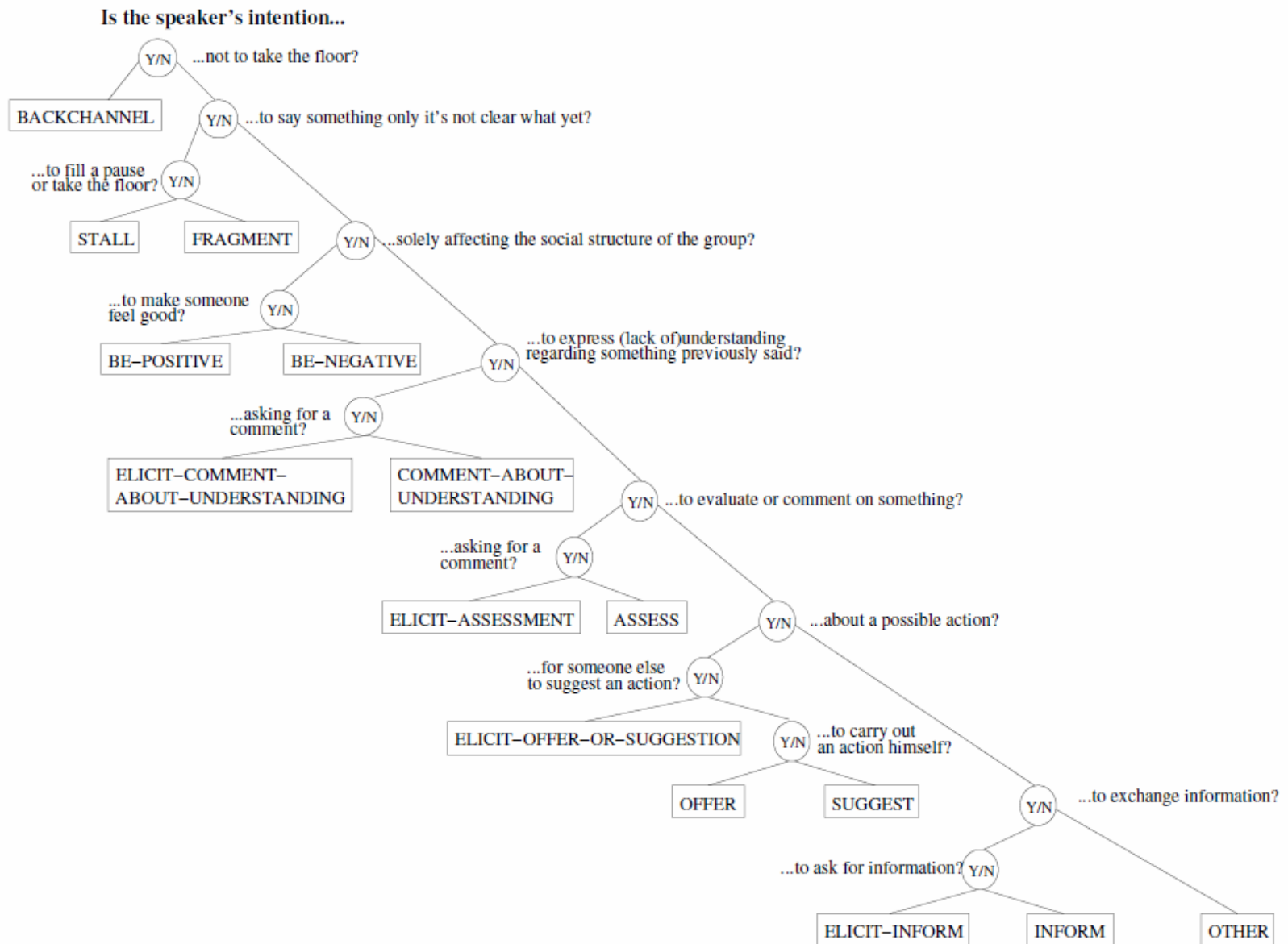
Label	Occ.	Label	Occ.	Label	Occ.
B	15,180	S	51,304	S^RP	7,612
H	12,288	S^AT	8,280	S^RP^DO	41
Q	5,320	S^AT^RI	273	S^RP^RI	436
Q^AT	3,137	S^DO	3,935	S^RN	2,219
Q^AT^RI	69	S^DO^RI	32	S^RN^DO	38
Q^DO	239	S^PO	791	S^RN^RI	46
Q^RI	60	S^RI	765	S^RU	1,298
				S^RU^PO	61

AMI corpus and DA tagset

- Ca. 57,000 utterances tagged with DAs (2006)
 - 72 4-people 1-hour meetings (<http://corpus.amiproject.org>)

 - “Pick exactly one label from the following ones”
 2. Acts about information exchange: INFORM, ELICIT-INFORM
 3. Acts about possible actions: SUGGEST, OFFER, ELICIT-OFFER-OR-SUGGESTION
 4. Commenting on Previous Discussion: ASSESS, COMMENT-ABOUT-UNDERSTANDING, ELICIT-ASSESSMENT, ELICIT-COMMENT-ABOUT-UNDERSTANDING
 5. Social acts: BE-POSITIVE, BE-NEGATIVE
 6. The bucket class: OTHER
 1. Special classes to allow complete annotation (*not really DAs*): BACKCHANNEL, STALL, FRAGMENT
-

AMI decision tree for DA annotation



“DIT Multidimensional annotation using DA types” (H. Bunt et al. 2006)

- For each utterance at most 1 *dialogue act type* per dimension
 - Dialogue act type:
 - either a general-purpose communicative function name + a dimension name, e.g. **WH-Question Auto-Feedback** (“*What did you say?*”)
 - or just a dimension-specific communicative function name, e.g. **Apology** (“*I’m sorry*”)
-

“Multidimensional annotation with the DIT++ scheme” (H. Bunt et al. 2006)

- Generic scheme from DIT incorporating elements from (SWBD-) DAMSL, TRAINS, GOT-IM, and other schemes, with open slots (e.g. for discourse rels)
- 12 dimensions
- In total approx. 95 communicative functions:
 - 42+ general-purpose
 - 51 dimension-specific):
 - 20 feedback
 - 21 interaction management
 - 10 social obligations management
- Applied to AMI multiparty dialogues (O. Petukhova)
- Being applied to human-human and human-computer dialogues in Dutch, Spanish and Italian

To conclude:

What could be the object of an ISO standard for dialogue acts?

(apart from segmentation issues)

A tagset, or rules for defining tagsets?

1. A tagset: can we find “the most general” one?
 - given the variety of purposes, theoretical approaches and data, this seems difficult
 - finding conversion rules to/from all tagsets to the “pivot” one seems difficult as well
 2. Rules for defining tagsets: a more promising approach?
 - e.g. ISO 16642, Terminological Markup Framework
 - XML “metaformat” of which actual terminological formats were instances (e.g. MARTIF et GENETER formats)
 - e.g. ISO/IEC 9126 & 14598, Software Evaluation
 - principles for designing quality models & evaluations + examples
-

A source of inspiration: H. Bunt @ ACL-SIGSEM 10-11 Jan 2005 meeting

- Definition 5: A layered multidimensional dialogue act assignment system is a 7-tuple $A = \langle GP; DS; D; f; C; L; T \rangle$, where:
 - GP is a set of general-purpose communicative function names;
 - DS is a taxonomy of dimension-specific communicative function names;
 - D is the taxonomy of dimension names that mirrors the DS taxonomy, extended at the top with dimension name 'task/domain';
 - L is a set of layers (i.e., set of (sets of...) dimensions of D);
 - f, C and T are as before [f is a function assigning tags to utterances (which may be simple elements of D, or complex structures built from D elements)] [C is a set of constraints on admissible combinations of tags] [T is a set of additional labels that f may assign to utterances (T contains such labels as inaudible and abandoned)], except that f is now a function from utterances to sets of tags (or labels from T), each tag being either an element from DS or a pair $\langle g; d \rangle$ with $g \in GP$ and $d \in D$.
 - ***Inspired by DIT++, but potentially very general***
-

Note on LIRICS D4.2 Data Categories

1. Do they correspond to the DIT++ tagset?
 - then maybe other tagsets need to use other categories

 2. Or do they describe “universal categories” of DAs?
 - are we sure such categories exists?
 - are **names/terms** universally accepted? are their **definitions** universal?
 - do categories change radically with the theoretical framework?
 - e.g. “speech acts”: widespread & still controversial...

 3. A prudent proposal
 - consider that this is a non normative, but generic example
 - good generality
 - need constraints for application: one or more tags per utterance
-

Items to include in ISO standard on DAs

- “For a DA tagset to conform to the ISO standard, the following aspects must be specified:”
 1. Objective of tagset, NLP application, domain
 2. Theoretical grounding
 3. List of tags, organized by subsets [“dimensions”]
 - plus: structure, hierarchies, default tags, nb. of combinations
 4. Definition of each tag and its conditions of application
 5. *A priori* constraints within each subset (mutually exclusive tags or not) and across subsets (if any)
 6. Existing annotated data, reliability figures, tools
 7. Conversions to/fro with other tagsets
-

Maintain appropriate level of generality

- Define all the terms appearing in the previous items as generally as possible
- State each requirement precisely and give examples of possible ways to answer it
- Applications: non normative Annex
 - describe “in ISO terms” some existing DA tagsets
 - use DIT++ as a prototypical example