Language-acquisition Inspired Sustainability Modelling For Application Profiles

Dr Emma TONKIN
University of Bristol
e.tonkin@bristol.ac.uk

DC-2015 – São Paulo, Brazil
Application profiles

* Pragmatic approach: contextualised application of standard approaches
* Reflect interdisciplinary boundaries; artefacts of situated practice
* What about sustainability?
  * External/internal political,strategic mandates
  * Staff turnover; changes in organisational structure; changes in social attitudes?
Types of change

* Linguistics: Semantic evolution, shift, drift, change...
* Semantic evolution: change in some part of a system affecting how a term/concept is understood
* Known to be destabilising factor in software ontologies: necessary to recontextualise items not in active use, potentially a disruptive process
Types of change (II)

* Semantic drift: ‘gradual change of semantic value as understood by the relevant community’ (Gulla et al, 2010)
  * **Intrinsic**: change wrt other concepts within frame of reference
  * **Extrinsic**: change wrt real-world referent
* Semantic change: *broadening, narrowing*
* **Dimensions of change** (Baruzzo et al): informational domain; technological domain; social domain
Fine. Change happens. So what?

Relevant questions:
- How fast will it happen?
- How severely will this affect us?
- How can we track it?
- How can we capture it?

Standard models: conceptual models; agent-based models; evolutionary theory; probabilistic approaches; ontological modelling; goal-oriented stakeholder modelling; cognitive modeling...

‘All models are wrong, but some are useful’ – Box (1987)
Simple model based loosely on Niyogi (2006); Kaplan et al (2008). In brief:

- Linguistic knowledge & behaviour in knowledge management = formal system.
- Humans hold a range H of systems
- To learn $H_n$ one requires exposure to a competent speaker of $H_n$
- One must have opportunity to learn all $n$ terms within system
In practice, not all terms are easily learnable: some concepts are hard to grasp.

In practice, it’s even harder: variation exists within any system!

And on the other hand, people can learn from documentation, although the accessibility of documentation may diminish over time (Kanhanbua, 2013).
Qualitative case study: continuous and discontinuous communities

Two sample cases:
- Close-knit team with low staff turnover and regularly-used application profile
- Ephemeral team produces AP, uses it, disbands; new team ‘inherits’ data, tries to make use of AP

Former case more likely in many museum/archive contexts valuing continuity of practice; latter more common in scientific research contexts

For simplicity we assume that learnability of all terms is total

Key factor: probability of learner coming across an instance of a term in use
Case 1: Highly connected team

- High probability at any time that a learner encounters a learning event
- Learner rapidly begins to learn terms
- Learner rapidly completes their study
- All is good!
Case 2: sparsely-connected/disconnected context

- If no learning events take place, new learners cannot become fluent
- No need to explicitly model this (it’s obvious)
- Instead we model low-frequency learning events
- Learning is still possible, but less likely, so fewer productive events occur, reducing learning rate relative to Case 1
Or, in short...
Learning rates and realism

- ‘Markov timesteps’ aren’t real life.
- But in real life, the S-shaped curve seen here is very real and appears often in study of language change.
- Problem is, it eventually becomes unrealistic that a learner completely learns any given system.
- So terms fall out of use, change inadvertently, or break.
What do we do?

- Would-be learners in difficult situations must rely on documentation.
- AP curators can track and even predict likelihood of change issues by evaluating AP usage patterns.
- To do that, we require data about usage, and about the communities in which the material is used: are they disconnected islands or thriving communities?
- Arresting change is difficult and potentially unrealistic. Evaluating and recording risk is much easier.
- We collect metadata about provenance. Should we not also collect metadata about change?
- Semantic change may be beneficial, too... but we still need to know it happens.
Community structure and history is of great significance in evaluating usage of specialist termsets

Evaluation of rate of change is strongly dependent on context(s) of use (and complexity of termset)

Where little experience is available, people may use low-fidelity learning strategies to ‘fill the gaps’

Open question: is provenance metadata sufficient for evaluating semantic change, or do we need to store more information about the relationships between terms, communities and records?