MODELLING CULTURAL VARIATIONS IN INTERPERSONAL COMMUNICATION FOR AUGMENTING USER GENERATED CONTENT

1

Dr Dhavalkumar Thakker (Dhaval)
Research Fellow
Artificial Intelligence Group
University of Leeds
Leeds, UK

ImREAL

http://www.imreal-project.eu/

OUTLINE

- ImREAL Project
  - Problem area in the project
  - Domain : Cultural variations in interpersonal communication
- AMOn+ ontology development
  - Methodology
  - Bottleneck & solution
- AMOn+ Evaluation
- Conclusions

THE IMREAL PROJECT

Linking simulated environments and real-world experience

Simulated environment

Real-world experience

Business simulation & Buddy program

Medical interview training

Gap
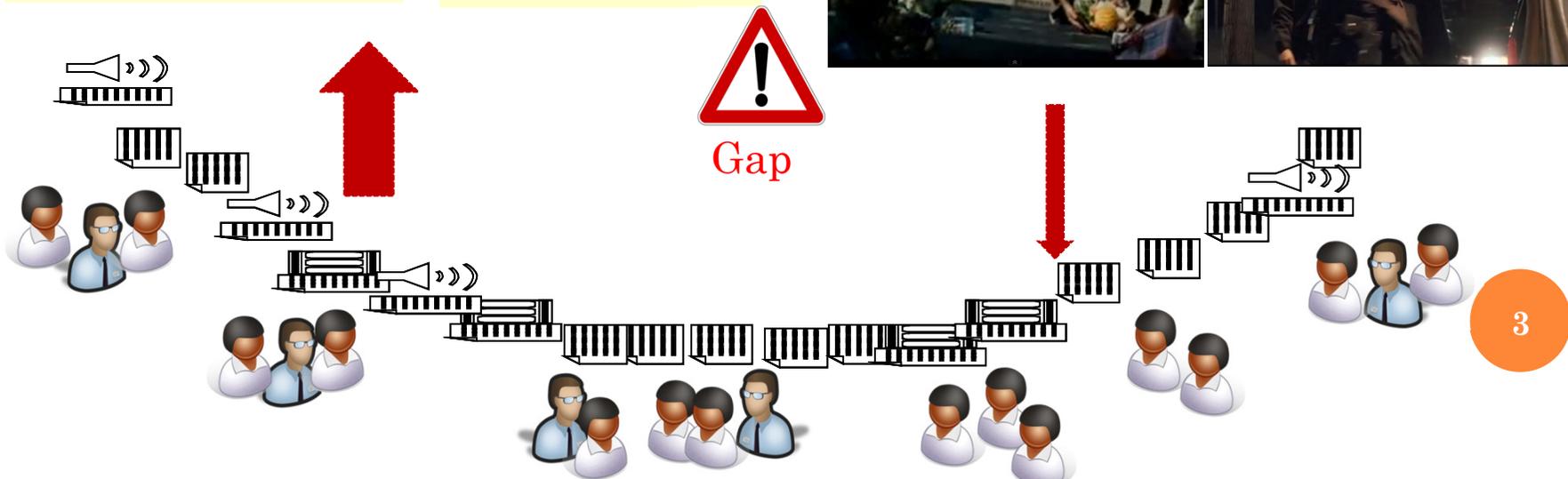
REAL-WORLD EXPERIENCE: DIGITAL TRACES ON SOCIAL WEB

Every day:
175 million tweets [average]

KleinEpstein&Parker @KEPRebels
What you wear at a **job interview** is just as important as the hand shake bit.ly/Nz8mbe #**interview** #**wear**
Collapse ← Reply ⇆ Retweet ★ Favorite

50% Maori/ 50% Tonga @Native834   18 S
Job **interview** at 10:30 today ... gotta keep reminding myself about **EYE CONTACT** ... i dont really look at people when they're talking
Expand

Him with the part.. @nonrappin_nas   12 S
The manager stumbled through the **interview**. Didn't maintain **eye contact** or anything. Like she was scared.
Expand

Every day:
2 million blog posts [average]

Sharing My Job Interview Experience
Posted by accountingweb on Feb 9 2011 0 3697

Sharing My Job Interview Experience

By Yu Li

Member of IMA's Young Professional Committee

Student at George Washington University

It is said that an awesome résumé gets you an interview, and a successful interview gets you an offer. Indeed, interviewing is an important skill. It involves not only expressing yourself clearly, but also communicating it to your interviewer. As a current graduate student, I have had several interview experiences. There are errors and little shining points in those experiences that I would like to share with you now.

Know Yourself

About two years ago, I had a phone interview with a big local accounting firm for a tax internship opportunity. Because this was my first formal interview, I did a lot of homework. I checked through the possible interview questions and prepared answers, while also researching the company's background. I found the WikiJob website was quite useful for me during my preparation.

It took me a whole day to prepare myself. Even so, my voice was still shaking when the interview began. "Tell me about yourself in one minute. Give me an example of yourself as a team leader." I answered these questions without hesitation because I already had them on my cheat sheet. Then the interviewer asked, "Why would you like to join our company?" I answered her question from both professional advantage and social responsibility of the company. She seemed to be satisfied with my performance so far.

"Great, now tell me about one of your biggest weakness." I didn't expect this question. I remembered clearly that it was among the questions list, but I unfortunately didn't prepare for this one. "Err... I am a ..." This is not an easy question, because these weaknesses may prevent the interviewer from hiring me. After a long pause, I decided to say what came to my mind first: "I am a quick person, when something isn't done in time, I would get worried and try to finish it as soon as possible."

Social media:
Source of digital traces with people's experiences in various situations

4

SEMANTIC AUGMENTATION OF DIGITAL TRACES IN CULTURE/IC DOMAIN

- Semantic Augmentation for aggregation and analysis of digital traces.

- Semantic augmentation is a process of attaching semantics (in the form of "concepts") to a selected part of a text to assist automatic interpretation of the meaning conveyed by the text.

- In order to semantically augment web content it is necessary to have a semantic model in the form of an ontology (Bontcheva & Cunningham, 2011).

- No ontology of this domain (cultural variations in interpersonal communication) exists.

DEFINITIONS

Broad notion of culture

Culture can be defined as a set of beliefs, values, behaviours and practices that characterise a **given group of people**

[Kashima, 2000]

Narrowed scope: National Culture

Nationality and countries have been used as fairly **reliable indicators** for tackling cultural diversity

[Gupta et al, 2002; Globe clusters; Hofstede dimensions]

Kashima, Y. Conceptions of Culture and Person for Psychology, *J. of Cross-Cultural Psychology*. 31, 1 (2000), 14-32.
Gupta, V., Hanges, P.J., Dorfman, P. Cultural clusters: methodology and findings. In Journal of World Business, 37,2 (2002) 11-15.

AMOn+   ACTIVITY MODEL ONTOLOGY

- Represents the main aspects of an interpersonal communication activity; indicates these aspects which have cultural variations.

- Ontology designed with socio-technical approach
  - Developed using ontology development methodologies
  - Based on social science theories
  - Based on Linked Data

- Evaluation
  - Fit for purpose evaluation
  - Using Gold Corpus with Information Retrieval evaluation techniques

AMON++ DEVELOPMENT

Using Uschold & King and METHONTOLOGY methodologies

Planning

- Social & Computer scientists
- Intercultural team

Scope & Purpose

- Requirements from ImREAL Project
- Culture only in the context of Interpersonal communication
- Cultural aspects for nationality

Ontology capture

- Identifying key concepts and relationships
- Squeezed approach - empirical (competency questions and other data) and theoretical (use of literature)
- Advising social scientists in preparation of Glossary of Terms (GT)
- A Priori Modularisation:
  - Core Ontology
    - Activity Module
    - Interpersonal Communication (IC) Module
    - Cultural Module

AMON++ DEVELOPMENT: SPECIFICATION

Domain:	**Cultural Variations in Interpersonal Communication**
Date:	01/09/2013
Conceptualised By:	Emmanuel Blenchard, Stan Karanasios, Dhavalkumar Thkker, Ronald Denaux, Vania Dimitrova
Implemented By:	Dhavalkumar Thkker, Ronald Denaux, Vania Dimitrova
Purpose:	Ontology underpinning for Semantic Augmentation services in the domain of cultural variations in interpersonal communication
Level of Formality:	Formalised (OWL)
Scope:	We restricted conceptualisation of cultural aspects related to interpersonal activity.
Sources of Knowledge:	Memetic Theory, Dual Inheritance Theory, Sperber's Epidemiology of Representation, Distribution of cultural conceptualizations, Culture and Cognition, System of Values of Hofstede, GLOBE system of values, Schwartz Value Inventory, Cultural Intelligence, Cultural framework of Alwood, Framework for intercultural training of Bennett, Research on specific cultural variations, Cultural Framework of Hall, Politeness Theory

Engeström, Y. (1987).
Activity Theory

10

AMON++ DEVELOPMENT: INTERPERSONAL COMMUNICATION MODULE
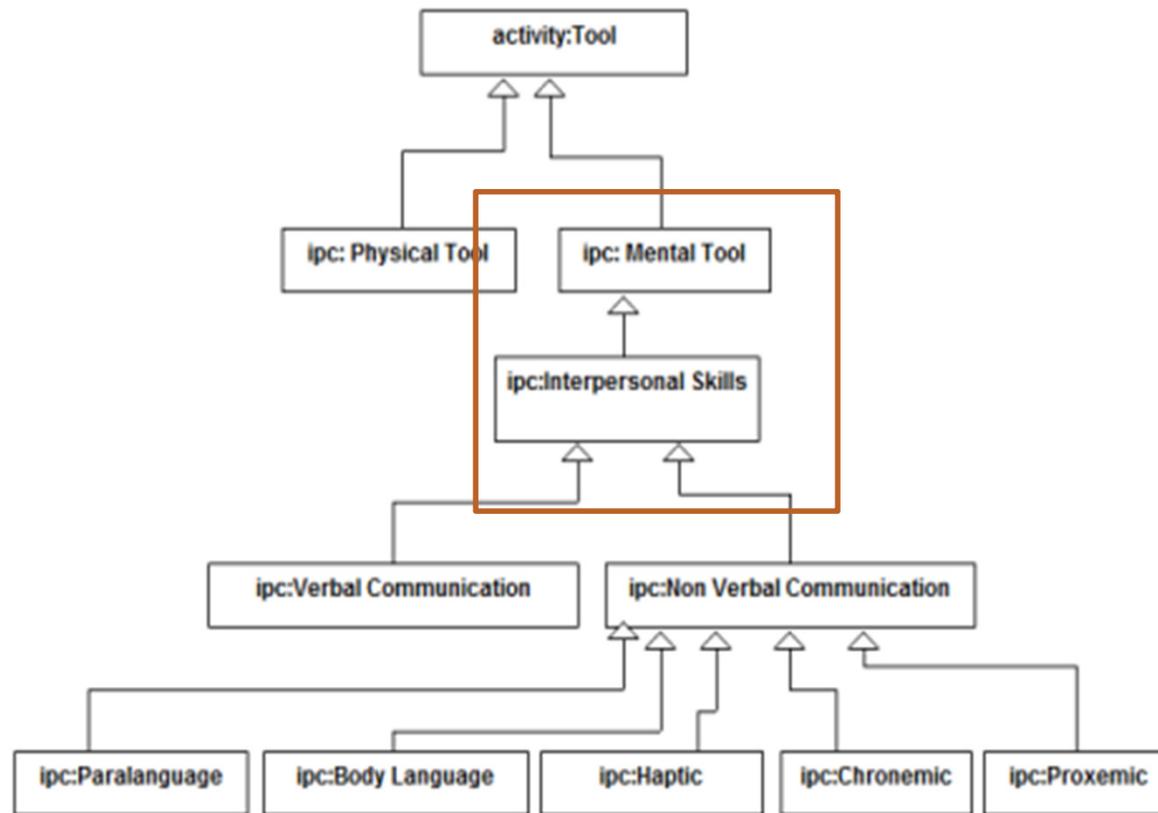
Extending tools into **interpersonal skills**

AMON++ DEVELOPMENT: IC MODULE
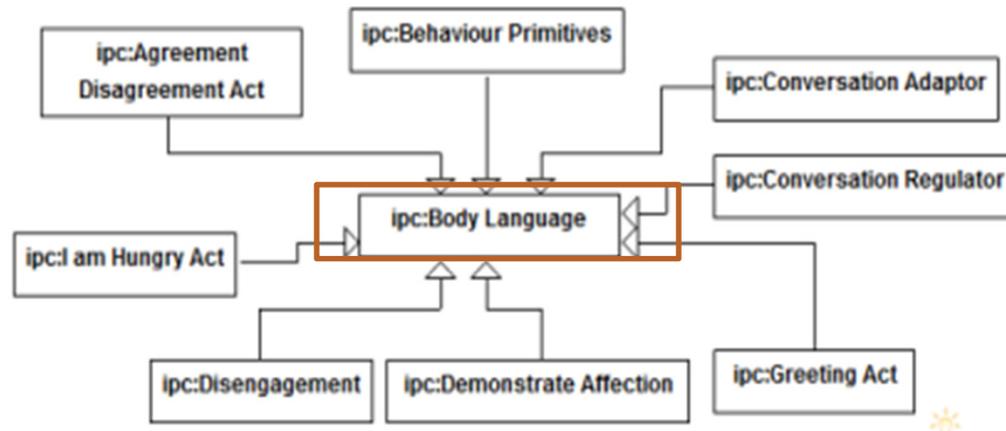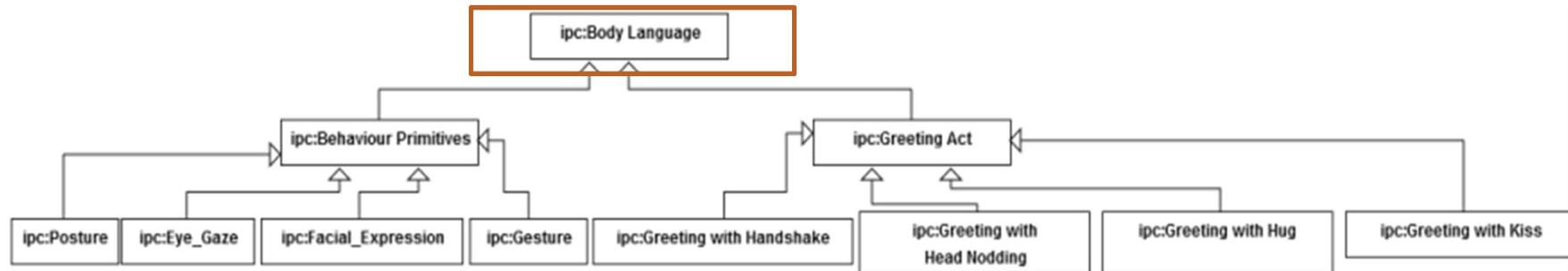
Concept of "Body Language"

13
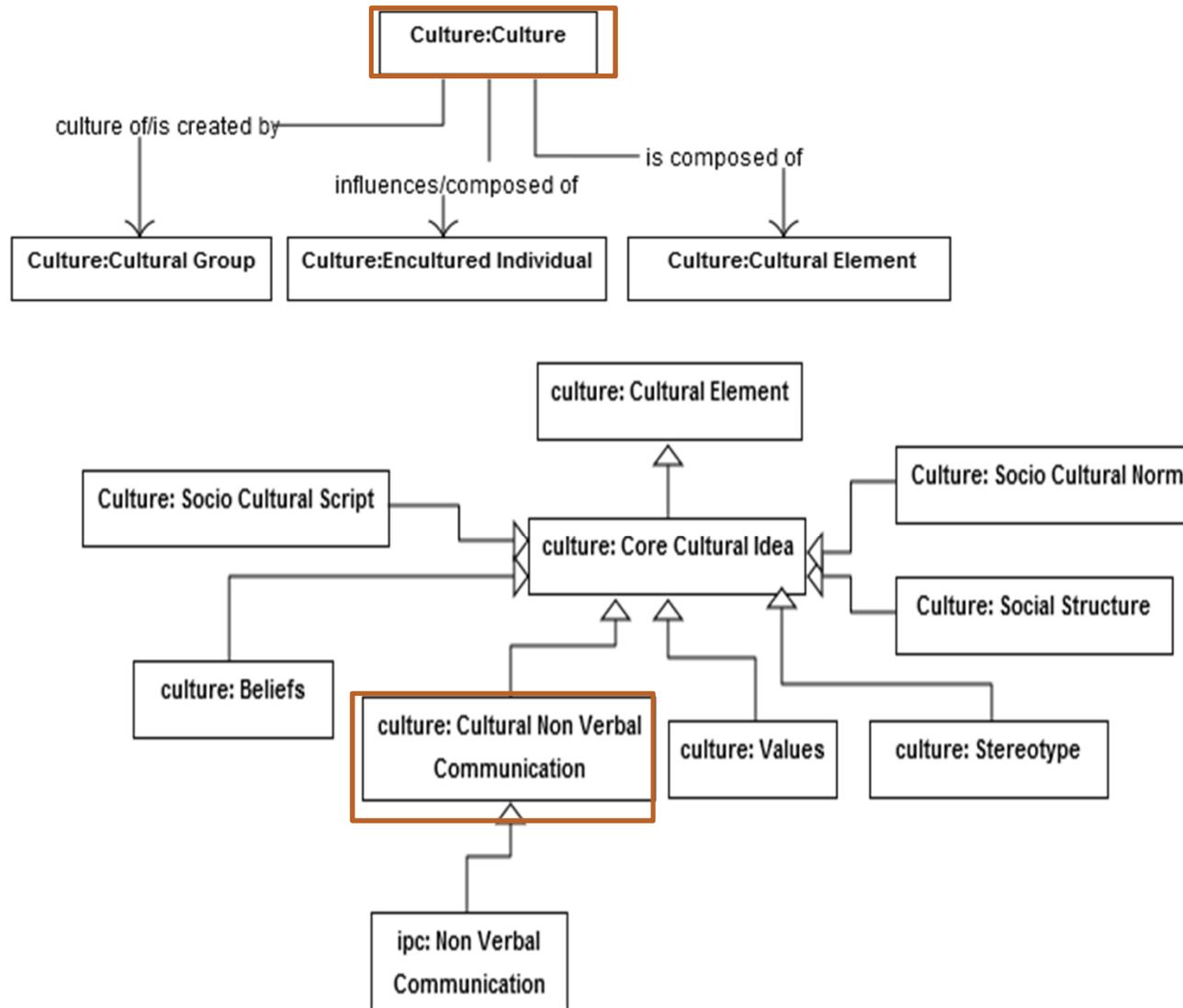
AMON++ DEVELOPMENT

Using Uschold & King and METHONTOLOGY methodology

Planning

- Social & Computer scientists
- Intercultural team

Scope & Purpose

- Requirements from ImREAL Project
- Culture only in the context of Interpersonal communication & social signals
- Cultural aspects for nationality

Ontology capture

- Identifying key concepts and relationships
- Squeezed approach - empirical (competency questions and other data) and theoretical (use of literature)
- Advising social scientists in preparation of Glossary of Terms (GT)

Ontology coding

- OWL Encoding of developed glossary- CORE ontology
- Using glossary format and natural language definitions
- Output: OWL Ontology. Available from: http://imash.leeds.ac.uk/ontologies/amon/

14

AMON++ DEVELOPMENT: ISSUES WITH THE ABSTRACT NATURE OF THE ONTOLOGY

Social content operates on concrete level
Abstract level – Greeting
Concrete level – kiss on cheek, handshake

Yes. In Brazil men also shake hands when they meet each other. Women kiss each other in Brazil too, but the number of kisses varies according to the region of the country (one, two or three kisses, depending on the state) For example, in Sao Paulo, they exchange one kiss only. In Brasilia, they usually say hello with two kisses. It does no matter whether the meeting is the first or not.

Reply

- While the literature provided key theoretical foundations, an apparent limitation is that the resulting core ontology describes cultural variations at an abstract level.
- For example, there are no concrete instantiation for 102 classes out of possible 125 classes (81.6%) from the ontology.

- Core Ontology providing upper level terms for enrichment
- Enrichment using Linked Data/DBpedia

15

EXPLOITING RICHNESS IN DBPEDIA CATEGORISATION

Categories
(e.g. Human_positions
Nonverbal_communication)

Other categories
(e.g. Gestures)

Other categories
(e.g. Greetings)

Level 1

Level 2

Level 3

Resource/Page
(e.g. "Body Language")"

Other Resources
(e.g. "Bowing")"

Resources
(e.g. "Cheek Kissing")"
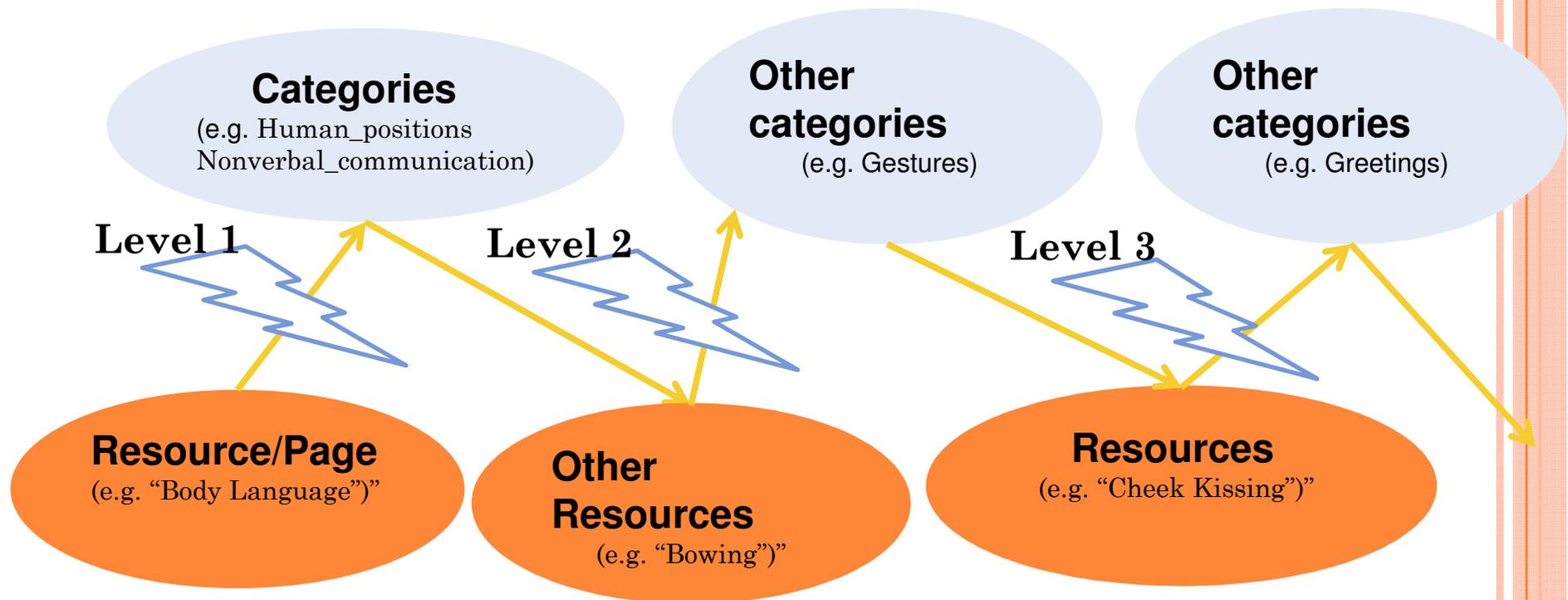
Category chains – different levels in DBpedia

16

ENRICHING AMON+ WITH DBPEDIA:
STEP 1: LINKING AMON+ WITH DBPEDIA

Mapping Mechanism

- 76 resources matches were found out of total 125 Classes(60.8%).
- At instance level, 40 out of total 22 matches were found (64.5%).
- In total, out of 187 entities (classes and instances), 116 matches (62%) were found from DBpedia using the mapping process described above.

17

HOW ABOUT MISSED MAPPINGS?

- **Type A: Specific or General Concept not Available in DBpedia**
  - The core ontology contains some concepts that are too specific and not used in routine interactions that a crowd sourced knowledge source such as Wikipedia is going to provide. E.g. Socio Culture Norm, Cultural Group Cohesion

- **Type B: General Concept Available but Specific Concept Not Available in DBpedia**
  - The core ontology contains some concepts that are too specific forms of some routine concepts that can be found in a crowd sourced knowledge source such as Wikipedia is going to provide. For example "Situation related Dressing"

- **Type C: Context Mismatch between AMOn+ and DBpedia**
  - This represents cases when contextual mismatch was found between two datasets. For example concept of "Operation"

EXPLOITING RICHNESS IN DBPEDIA: STEP 2: EXTRACTION

Extraction pipeline
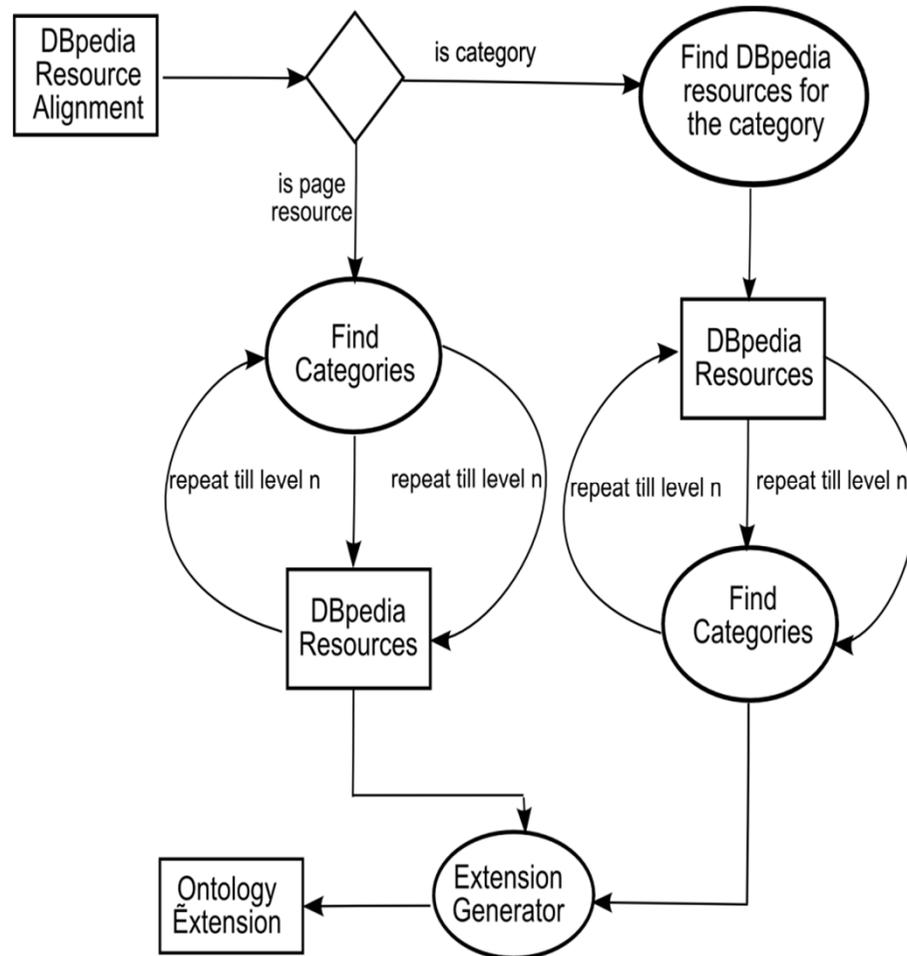
Level	Size (MB)
Level 1	1.1
Level 2	126
Level 3	1.27 GB

AMON+ ONTOLOGY SUITE: CORE + DBPEDIA ENRICHED

AMOn+ http://imash.leeds.ac.uk/ontologies/amon/
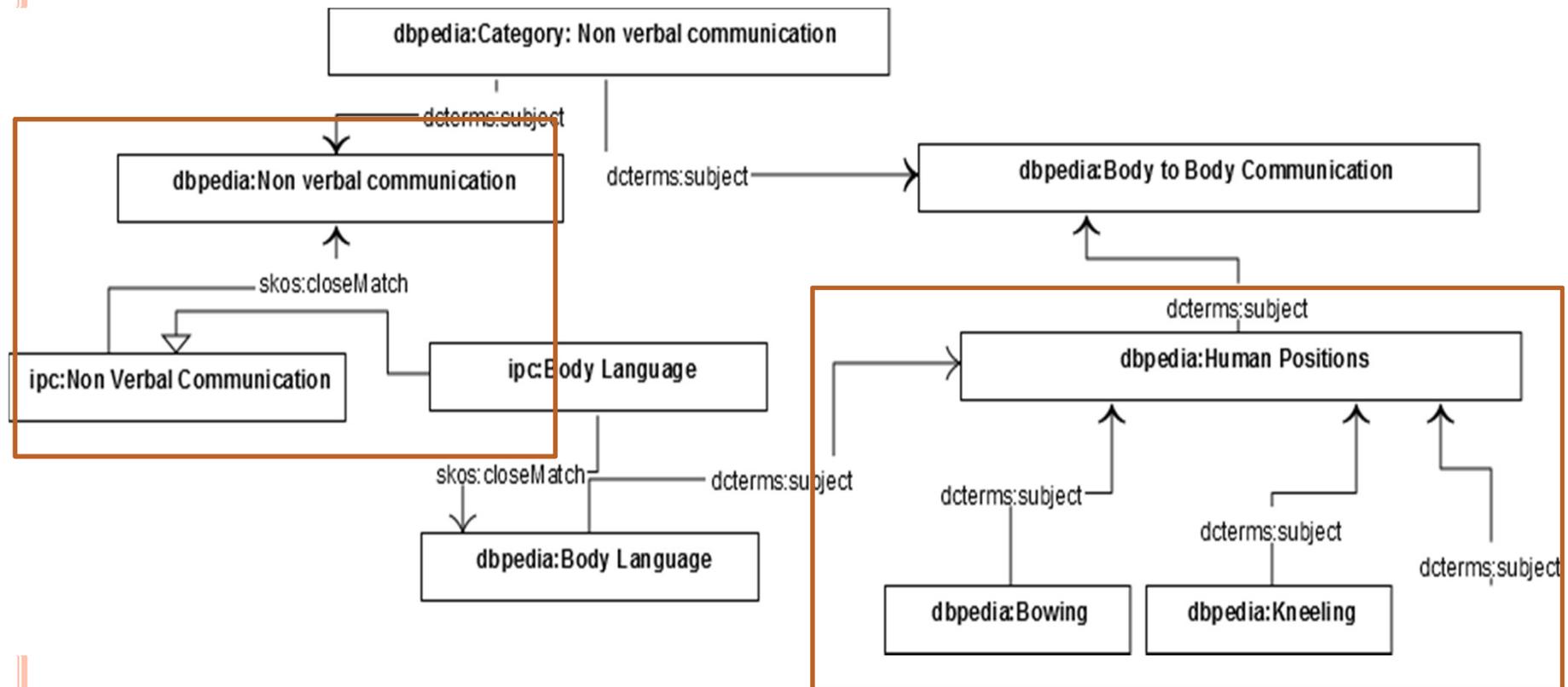
EVALUATION

Goal: To check whether it is fit-for-purpose for automated semantic augmentation of content containing cultural and interpersonal communication aspects

- It is acknowledged that for the applications that use ontologies for semantic augmentation and natural language processing tasks, **domain coverage is more important than the logical correctness to measure their fit-for-purpose utility** (Suárez-Figueroa, Gómez-Pérez, Motta & Gangemi, 2012).
- Domain coverage is a measurement of the extent to which an ontology covers a considered domain.
- To measure domain coverage, semantic annotations services using the ontology should match a very concrete **domain defined by an annotated textual corpus** (Brewster, Alani, Dasmahapatra, Wilks, 2004), **often referred as gold standard** (Brewster, Ciravegna, Wilks, 2001).

21

CONSTRUCTION OF TEXTUAL CORPUS

- Gold Corpus consisting of heterogeneous sources of digital traces that discuss aspects related to cultural variations in interpersonal communication:

  - **(a) User Generated Content** which are generally contains short text;

  - **(b) News Articles** that generally contain detailed story on a particular aspect. These types of resources generally have cultural aspects surrounded by a lot of context and

  - **(c) Wikipedia pages** on cultural variations in interpersonal communication which are very focused writings on cultural variations among different countries.

  - **(d) Specialised Web resources** which are similar to Wikipedia and are focused writings on cultural variations among different countries.

22

ANNOTATION OF THE TEXTUAL CORPUS

- **Sampled:** To sample the corpus to a size that is manageable for manual annotation, we have selected content from each of the content categories (i.e. Wikipedia, UGC, etc.) so that it covers at the least one country from the GLOBE societal clusters.

- **Annotation by:** An expert with Cultural Intelligence (CQ) score of 6.5 (out of possible 7) annotated textual content to identify key terms relevant to the domain as annotations. The Cultural Intelligence (CQ) score for the expert is measured using a method proposed by Van Dyne, Ang, & Koh, 2008.

23

ANNOTATION OF GOLD CORPUS

○ The expert was given the definitions of cultural variations in interpersonal communication and given the categories that the system focuses on. They were asked to read the comments and using GATE annotate all the terms they thought were related to culture variations in interpersonal communication and should be annotated by semantic annotation service using AMOn+.
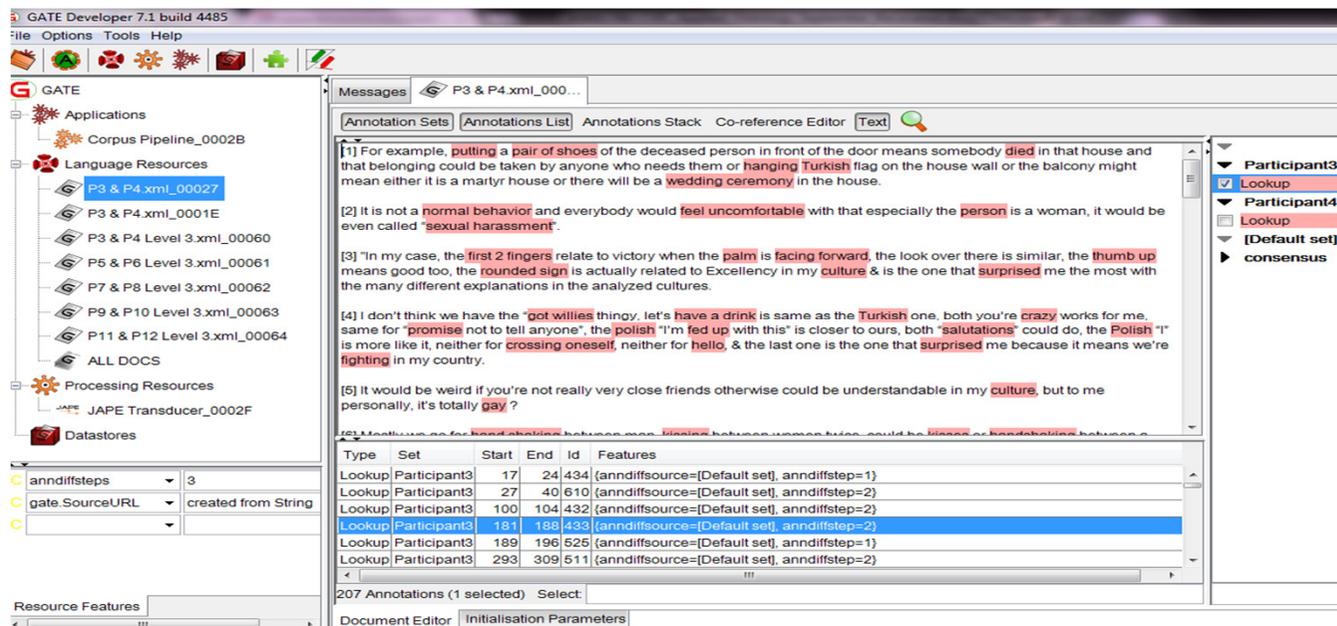
GATE Annotation Interface: gate.ac.uk

ANNOTATION OF GOLD CORPUS: INTER ANNOTATOR AGREEMENT

- To ensure the confidence in the annotation results, **we recruited another expert** who annotated a random sample from the first expert.

- A meeting was arranged between the two experts where the two experts agreed the conventions for annotation between them.

- The Inter Annotator Agreement (IAA) (Cohen, 1960) between two experts was **92.7%** which indicates **"Significant"** level of agreement between the two experts.

MEASURING COVERAGE: EXPERTS *AGAINST* SEMANTIC ANNOTATION USING AMON+

- Used GATE to annotate the same corpus using AMOn+

- The **classical information retrieval techniques** are suggested as alternative method to **measure the coverage of an ontology** while comparing against an annotated textual corpus – Precision, Recall and F-Measure

- To evaluate in this manner, **comparison of precision and recall of the service with human annotation of the textual corpus is required.** It is also important to trace the performance back to the presence and absence of concepts in the ontology.

- We followed this approach to measure coverage of AMOn+.

26

MEASURING COVERAGE: RESULTS

Precision
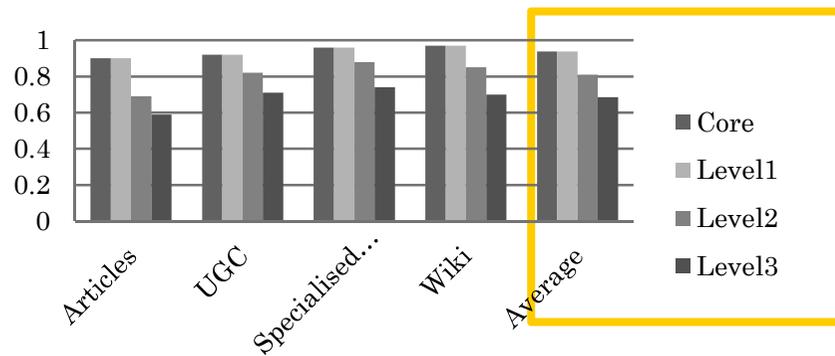
F-Measure
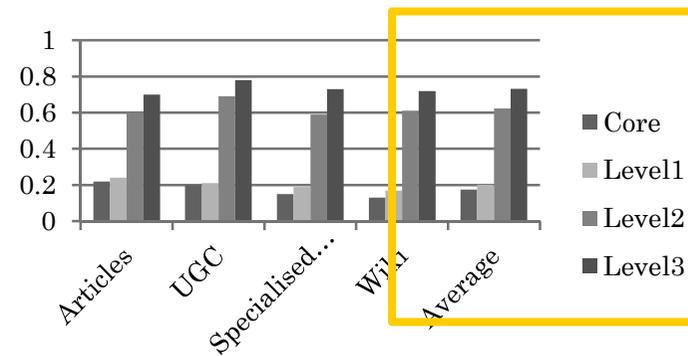
Recall
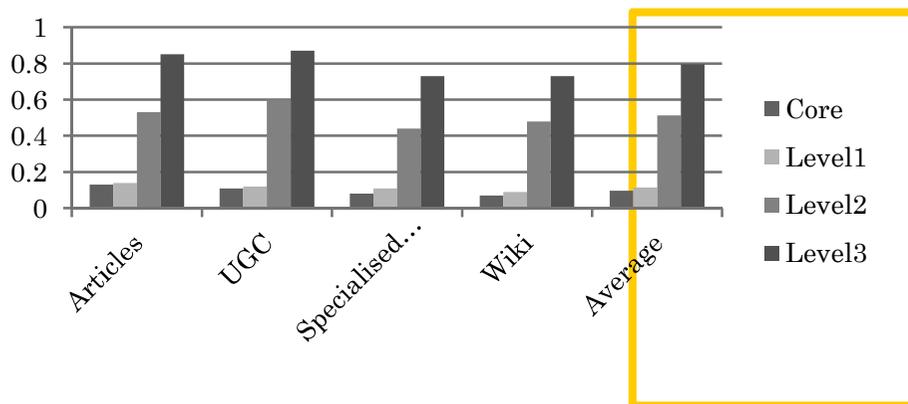
Level 3 gives
best performance

27

MEASURING COVERAGE: FALSE POSITIVE ANALYSIS

% of total terms missed
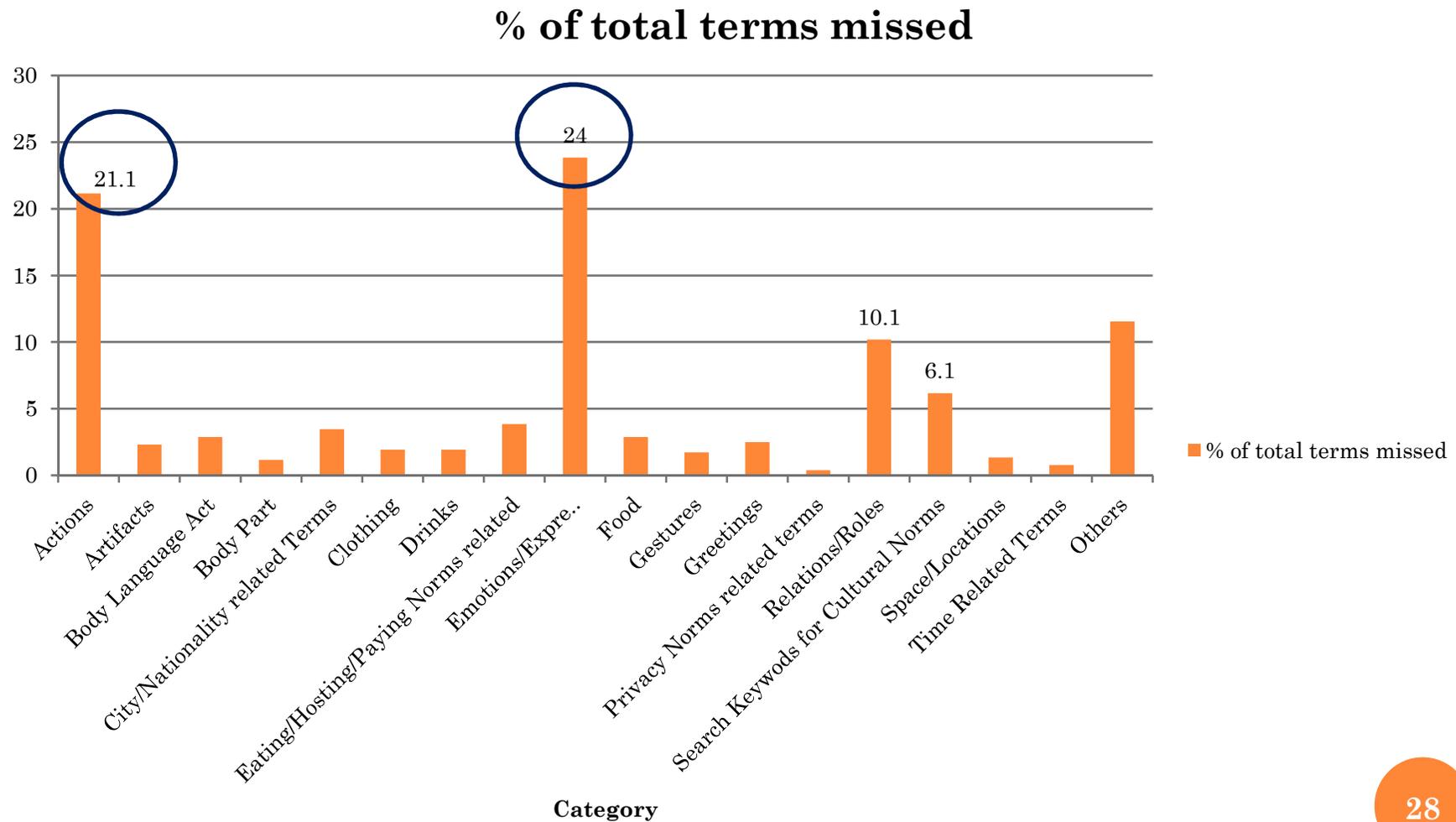
**Emotions
~24% of total**

**Actions
~21.1% of total**

MEASURING COVERAGE: FALSE NEGATIVE ANALYSIS

- **Organisations.** Some of the organisations have ambiguous surface forms (i.e. labels). Few examples of false negatives from the category of organisation are: dbpedia:Hands_On_USA which as surface form "hands on" and was one of the false negatives.

- **Names of People.** The concept of Person is part of the core ontology in AMOn+ with surface form "People" that is mapped to "dbpedia:Living_People" concept in DBpedia and in turn maps to living beings based on the categories pattern "YYYY births" or "YYYY deaths". For example, http://dbpedia.org/resource/Slimane_Raho retrieved using the category (http://dbpedia.org/resource/Category:1975_births).

- **Film/Programme/Song names.** AMOn+ contains concept of that is mapped to various categories of Films, Programmes and Songs from DBpedia. For example: "to be" (representing 2001 single http://dbpedia.org/resource/To_Be) and "who" (representing 1999 album http://dbpedia.org/resource/Who).

CONCLUSIONS
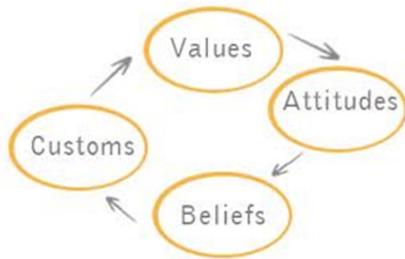
⇒ Culture is a key topic – simulators and beyond

⇒ Culture is a complex topic – scope & focus

⇒ Ontology of cultural variations in interpersonal communication

   ⇒ Relying on social theories for abstract concepts

   ⇒ DBpedia for extension and enrichment

⇒ DBpedia DOES know about culture (aspects) after all!

Mary Parker
Follett
(1868-1933)

Culture allows dealing with diversity

Unity, not uniformity, must be our aim. We attain unity only through variety. Differences must be integrated, not annihilated, not absorbed.