Machine Translation in CNGL II

Qun Liu
qliu@computing.dcu.ie
Dublin City University
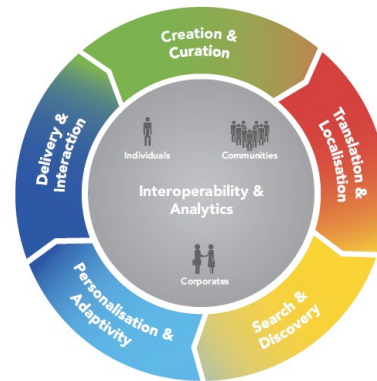**http://www.cngl.ie**

The CNGL Centre for Global Intelligent Content

The Centre for Next Generation Localisation (CNGL) is an Industry-Academia partnership funded jointly by Science Foundation Ireland (SFI) and industry partners. CNGL II (2013-2015) is the second round SFI foundation which aims to revolutionise the way people can seamlessly interact with content, systems and each other to achieve unprecedented levels of access, efficiency and empowerment. Through its Global Intelligent Content research and technology CNGL is creating breakthroughs in the creation, processing, unification and seamless integration of multilingual, multi-modal and multimedia content.

CNGL II Themes

- Creation & Curation
- Translation and Localisation
- Search & Discovery
- Personalisation & Adaptivity
- Delivery & Interaction
- Interoperability & Analytics

The themes C&C and T&L are directly related to machine translation.

Creation & Curation: Content With Purpose

The Creation and Curation theme turns raw data into information and knowledge, across languages, encompassing the full spectrum of potentially noisy user-generated and corporate content.

This theme provides the technologies to extract meaning and sentiment from text data and novel ways of visualising data and methodologies for aligning and connecting noisy meta-data across massively inter-linked intelligent content. Here, we transform un-structured content into (semi-) structured representations of information, adding core value to "raw" content and enabling later stages in the Global Content Value Chain.

The services provided by Creation and Curation underpin Intelligence in Global Content in two important respects:

1. content analytics extracts valuable information from data (turning data into intelligence)

2. imeta-data generation and annotations provide the foundations for Intelligent Content capable of being self-describing and aware of some of its core properties and effects.

Translation & Localisation: Communicate Globally

The Translation & Localisation theme will address quality improvement in translation technologies coupled with new approaches to localisation that maximise the value of human linguistic judgements across the widest range of domains.

This theme will explore new challenges in audio-visual localisation (games and affective content) and highly collaborative and community-based models, integrating state-of-the-art language and translation technologies. It will investigate the impact of language and translation technology on language usage in society and on language professionals.

This research will make important contributions to allowing content to transcend language barriers through: deployable methods for targeting and improving MT; effective localisation workflow and data sharing patterns tailored to the needs of crowd-sourcing and the translation of multimodal/multimedia content; and improved support for LT-assisted human translation skills in commercial and open-source tools sets.