

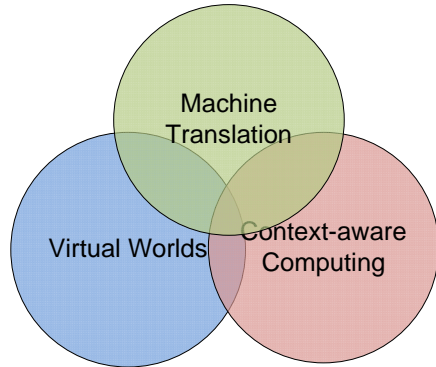
Ying Zhang<sup>1</sup>

Nguyen Bach<sup>2</sup>

<sup>1</sup>CMU Silicon Valley

<sup>2</sup>Language Technologies Institute

Carnegie Mellon University



- Virtual Worlds (VWs) such as Second Life are becoming increasingly popular for many interaction/collaboration activities.
- Just as in real world, **language barriers** exist VWs.
- We develop machine translation services to facilitate multilingual communication in VWs.
- Context** plays an important role in disambiguation and generate correct translations.
- Context information is easier to capture in VWs than in real world.
- We study **context-aware machine translation** in VWs.

## Virtual World Applications



Virtual Education 3d Reality Classes



Museum



Seminar and discussion



Sightseeing

## Machine Translation for Virtual Worlds



- Intercept incoming/outgoing messages and redirect to translation server.
- Phrase-based Statistical Machine Translation system based on CMU's SMT project.

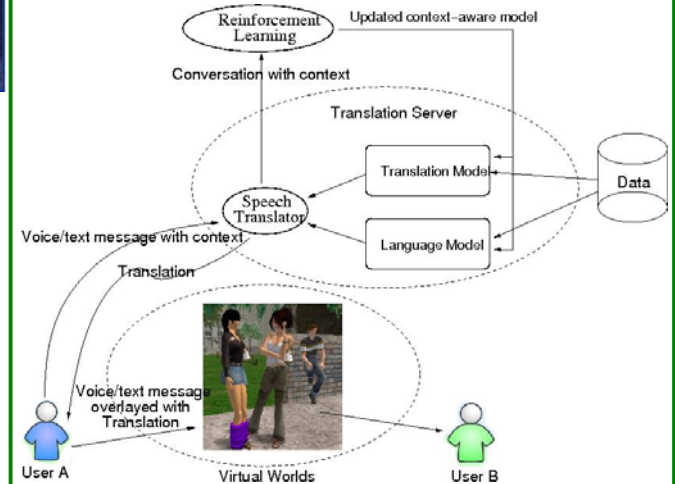
## Context-aware Machine Translation

Example 1	Chinese sentence: 在这儿能买到歌剧的票/tickets, admission, votes/ 吗?	Reference translation: Can I buy tickets for the opera here?	MT output: Here's opera can buy votes?
Example 2	Chinese sentence: 有红/red/ 的吗?	Reference translation: Do you have red wine?	MT output: There are red it?
Example 3	Chinese sentence: 下一班是几点/points, hours, dots/ ?	Reference translation: When is the next one (tight/train/bus/ferry) ?	MT output: The next is a few points?

Table 2: Examples of incorrect machine translations generated by Google's online MT system.

- Current MT systems look at surface form only
- Correct meaning of a sentence depends also on **non-verbal context information** such as time/location/users' gender/social background ...
- Conversational speech** tends to be more concise and more context-dependent.
- Context information is more accessible in VWs than in real world which requires physical sensors and sensing algorithms (GPS for location; accelerometers for gesture; gender?)

- Context-aware Language Model
  - Max-Entropy language model framework
  - Context feature serves as additional "knowledge source"
  - Logged conversational data with context information is used to train the model
- Context-aware Translation Model
  - Estimate  $P(f|e, C)$ : probability of  $f$  as  $e$ 's translation given context  $C$
- Training context-aware models via conversation analysis
  - Bootstrap the MT system using context-independent model
  - "What do you mean", "I don't understand you", and *walk-away* indicate that conversation failed and translation is bad
  - Adjust model probabilities based on feedback from automatic conversation analysis



- ### Ongoing Work
- Improving translation service performance and robustness
  - More conversational data collection
  - Analyzing context information
  - Compare context-independent MT vs. context-aware MT