Challenges for ICALL

Cornelia Tschichold

Swansea University

What is ICALL?

- corrective feedback in the context of
  - written L2 exercises (constrained text)
  - written L2 free-form text
- using NLP
- ICALL is sometimes also used to include system that
  - give corrective feedback on pronunciation (Engwall (2012)
  - give individualized feedback
  - use chatbots (e.g. Jia 2009), microworlds, etc. in a CALL context

A bit of history

- Early ICALL helped to produce exaggerated expectations
  - systems that can do what teachers do
  - much more cheaply
  - much more quickly

blue sky thinking

- Expectations could not be fulfilled, leading to
  - a backlash, and
  - mainstream CALL moving its focus to CMC

"the promise of intelligent tutoring systems has never quite been realised on any significant scale" (Rushby 2013:52)

Challenges today

- need to understand that
  - (I)CALL cannot replace a teacher,
  - but **can** be very useful
  - ICALL can treat a **subset** of L2 errors only
    - some pragmatics/style and semantics errors are probably out of reach for ICALL
    - agreement, articles and collocation errors have been explored
    - other areas still to be explored

Two main approaches

rule-based

- best done on the basis of error analysis of a large error corpus
- requires formalizable mal-rules and relaxing constraints on grammars
- may be difficult to scale up

statistics-based

- sees errors as divergences from common language
- requires (very) large corpora, both for L1 and learner language
- may be difficult to customize

Problems with statistical approaches

- typically based on bigrams
- assumption that words are normally distributed
  - words are clearly not normally distributed

Word distribution

- Zipf's law
  - A very small number of (function) words are extremely frequent.
  - The most frequent 1000 words give a reasonable coverage of many texts.
  - Many words beyond the 10K range may never be encountered by a language user.
- Clearly, word frequency is an important tool for materials developers.

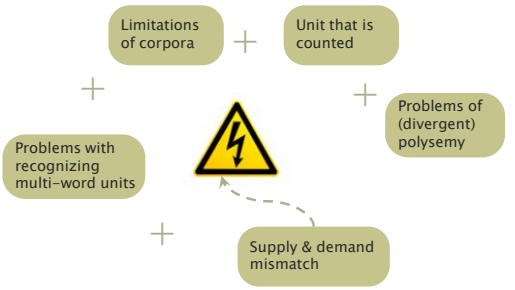
The problem with word frequency

- Vocabulary tests exploit corpus-based frequency lists
- Frequency lists can only come from corpora.
- Corpora are not unproblematic
  - size
  - coverage
  - bias

What counts as one item?

- orthographic word
- lemma
  - inflected forms only
- word family
  - inflected and derived forms
  - on an expansion scale

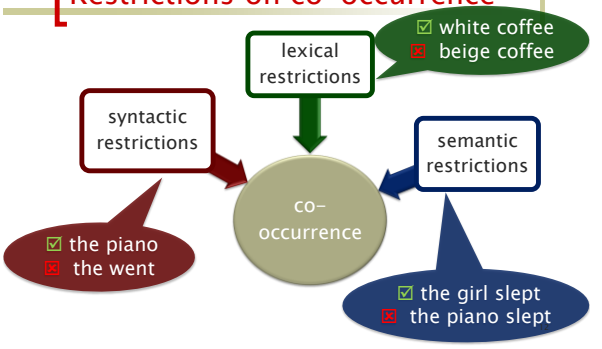
A vicious circle



Problems with statistical approaches

- typically based on bigrams
- assumption that words are normally distributed
  - words are clearly not normally distributed
- assumption that words are statistically independent
  - words are clearly not independent (**co-occurrence restrictions**)

Restrictions on co-occurrence



Challenges today

- vocabulary is more complex than typically assumed by (applied) linguists and CL
- need to understand that
  - ICALL can treat a **subset** of L2 errors only
    - agreement, articles and collocation errors have been explored
    - other areas still to be explored
  - we need standardized annotation schemes and test corpora
  - we need better comparability between systems (based on standard annotation and test corpora)

Comparing existing systems

- Annotation of errors is still a problem
  - Most systems use their own annotation system
  - S. Granger et al. on ICLE relatively well-known, but not a standard
  - Meurers
- Lack of standardized annotation leads to problems when comparing different systems.

Error detection and correction

- What do we want to correct?
  - grammaticality vs. acceptability
  - "mistake" vs. "error"
- comparison to correct version
  - more than one correct version possible
- What kind of feedback is useful to the learner?
  - at the stage they are at
  - in terms of their insight of the L2 system

The Common European Framework of Reference (CEFR)

- widely accepted and influential
- 6 steps

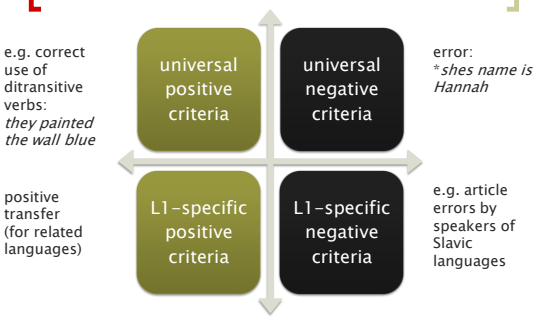
proficient
- C2 Mastery
- C1 Effective Operational Proficiency

independent
- B2 Vantage
- B1 Threshold — "can express opinions"

basic user
- A2 Waystage
- A1 Breakthrough

- but no information on typical errors

Criteria



What to flag at which stage?

- Any systems with a high ratio of over-flagging is problematic, but especially so at beginner's level.
  - Learners cannot be expected to discriminate between correctly flagged errors and over-flagging.
  - Precision therefore has to be as close to 100% as possible, even if this is at the expense of recall.
- At the beginners' level, only major errors should be flagged
- but what is best later on?

Challenges today

"We focus too much on the technology and not enough on the learning." (Rushby 2013:53)

- need to understand that
  - ICALL can treat a **subset** of L2 errors only
    - agreement, articles and collocation errors have been explored
    - other areas still to be explored
  - we need standardized annotation schemes and test corpora
  - we need better comparability between systems (based on standard annotation and test corpora)
  - we need a **clear focus** on the learner
  - ICALL has a large potential if used properly

What happens too often

- Many small and possibly very interesting projects vanish in a dead end because of
  - lack of funding and longer-term perspective
  - lack of insight into pedagogical needs
  - lack of uptake
- e.g. "ESL Assistant"

"The majority of exciting projects using handheld devices and mobile communications wither and die when their funding comes to an end." (Rushby 2013:54)

Concluding remarks

- We do not know how effective (especially long-term) today's systems are compared to human teachers.
- potential of word choice error correction largely unexplored
- L1-specific errors largely unexplored in the statistics-based systems

"This perhaps points to a fundamental difference between the goals of those who build automated error correction systems and those who educate second language learners." (Leacock et al. 2010: 100)