Report on

ERRARE'13: Errors In Natural Language Processing

November, 21-22 2013 - Ermenonville

The Errare workshop, a joint venture of the LIMSI/CNRS[1] and the IMMI[2] (Institute for Multilingual & Multimedia Information) took place on November 21-22, 2013 in the castle of Ermenonville, 60km north of Paris. The general topic of this interdisciplinary workshop was on "Error Analysis in speech and language processing"; it gathered, during two full days, approximately 50 scientists from diverse areas such as Automatic Speech Recognition, Multimedia Indexation, Spell-Checking and Machine Translation. A full program is provided in appendix.

The question of the automatic detection of errors, their analysis and categorization, in relationship with error prediction and evaluation is of prime importance for Machine Translation, and a number of talks specifically focusing MT were given. A first introductory talk on the issue was delivered by Lucia Specia (U. Sheffield), which provided a very thorough overview of the field of Automatic Metrics and Quality Estimation for MT.

M. Popovic (DFKI, Berlin) focused on the automatic categorization of error types in the context of High Quality MT; in particular she reported recent attempts at designing and using a multi-dimensional quality metric that were performed in the context of the European project QT Launchpad.

C. Kobus (Systran, Paris) presented a fine-grained analysis of post-editors' actions in the context of an industrial MT+post-edition workflow: looking at the actual changes that take place during post-edition provides an alternative way to categorize and diagnose machine translation errors. This makes the full automation of the full process, which remains itself error prone, even more desirable.

The talk by M. Simard (NRC, Ottawa) also looked at post-edition, but from a different perspective: as has proposed by the NRC team, post-edition data can effectively be used to train systems that can automatically correct the output of MT systems. As it turns out, these systems are quite error prone, and by analyzing the errors they commit, the speaker was able to suggest several ways in which such automatic post-edition systems could be improved.

Finally, N. Kübler (U. Denis Diderot, Paris), G. Wisniewski (LIMSI/CNRS) and F. Yvon (LIMSI/CNRS) finally exposed their views on the possibility of performing automatic

---

evaluations of Human translations, based on the findings of the MeLLange European project.

The complete proceedings (slides plus full video of the talks and QA sessions) will be shortly be available online on the workshop site:

http://www.immi-labs.org/spip.php?article227&lang=en

Program Workshop Errare 2013, 21-22 November 2013

20th of November (Wednesday)
Late afternoon: arrival of all the participants
20:30 dinner at the castle

Day 1, 21st of November (Thursday)

8:30-8:45	**Welcome/presentation of the workshop**
8:45 session 1	**Modelisation of Errors by machine**
8:45-9:40	**Keynote Julia Hirschberg,** Svetlana Stoyanchev, Philipp Salletmayer, Jingbo Yang, Alex Liu, Eli Pincus: "Identifying and Responding to Errors in Spoken Dialogue Systems"
9:40-10:05	Maja Popovic: "Automatic error analysis of machine translations"
10:05-10:30	Martin Sundermeyer, Volker Steinbiss, Hermann Ney: "Direct Observation of Pruning Errors - A Search Analysis Tool"
10:30-10:55	Atsunori Ogawa, Takaaki Hori, Atsushi Nakamura and Takanobu Oba: "Recognition rate estimation based on error type classification and its applications"
10:55-11:15	**Discussion**
11:15-12:30	Visit of Jean-Jacques Rousseau Park
12:30-13:30	*Lunch*
13:45 session 2	**Errors from multiple sources, complex errors**
13:45-14:40	**Keynote Josef Kittler:** "To fuse or not to fuse"
14:40-15:05	Juliette Kahn, Ilya Oparin, Olivier Galibert: "Evaluation is not just a score"
15:05-15:30	
15:30-15:55	Mirjam Ernestus: "Errors induced by pronunciation variants typical of casual speech"
15:55-16:15	**Discussion**
16:15-16:30	*Break*
16:30 session 3	**Impact of errors**
16:30-17:25	**Keynote Lucia Specia** "Is this translation fit for purpose? Predicting quality versus predicting errors"
17:25-17:50	Daniel Luzzati et al. "Human annotation of ASR error regions: Is 'gravity' a sharable concept for human annotators?"
17:50-18:15	Catherine Kobus et Jean Senellart "Machine translation error categorization by post-editions automatic analysis"
18:15-18:40	Benoit Favre, Kyla Cheung, Siavash Kazemian, Adam Lee, Yang Liu, Cosmin Munteanu, Ani Nenkova, Dennis Ochei, Gerald Penn, Stephen Tratz, Clare Voss, Frauke Zeller: "Automatic Human Utility Evaluation of ASR Systems: Does WER Really Predict Performance?"
18:40-19:00	**Discussion**
20:00	*Concert* **Benoît SAROCCHI and musicians** *traditional Corsican songs*
20h40	*Gala Dinner*
22h30	*concert (2nd part)*

Day 2, 22nd of November (Friday)

9:00 session 4	**Errors by machines and humans**
9:00-9:55	**Keynote Sarah Hawkins** "Illusions, contexts and domains of analysis in assessing errors in speech processing"
9:55-10:20	Michel Simard, Pierre Isabelle, George Foster: "Error Analysis of Error Corrections in Machine Translation"
10:20-10:45	Martin Cooke, Maria Luisa Garcia Lecumberri, Mate Toth: "Noise-induced slips of the ear"
10:45-11:10	*Pause+ check out*
11:10-11:35	Odette Scharenborg: "Comparing human and automatic speech recognition"
11:35-12:00	Bernd T. Meyer: "What's the difference? Quantifying errors in human and automatic speech recognition"
12:00-12:20	**Discussion**
12:20-13:35	*Lunch*
13:35 session 5	**Error Analysis**
13:35-14:30	**Keynote Mark Liberman:** "Machine learning from machine learning errors: Work in progress on parsing and phonetic alignment"
14:30-14:55	Sebastian Stüker: "The Influence of Topic and Domain Specific Words on WER - And Can We Get the User in to Correct Them?"

14:55-15:20   François Yvon, Natalie Kübler, Guillaume Wisniewski: "Human Errors and Automatatic Errors in Machine Translations. What are the Differences? "

15:20-15:35   Break

15:35-16:00   Dominic Telaar, Mark Erhardt, Tanja Schultz: "Error blaming based on Decoding Output"
16:00-16:25   Martine Adda-Decker, Barbara Schuppler, Lori Lamel, Juan Andrés Morales-Cordovilla, Gilles Adda: "What we can learn from ASR errors about low-resourced languages: a case-study of Luxembourgish and Austrian"

16:25-16:45   **Discussion**

16:45-17:00   Break

17:00-17:30   **Final discussion**

18:00   Bus departure