Workpackage 2
Interactive Translation Prediction

V. Alabau, G. Sanchis-Trilles, F. Casacuberta, Universitat Politècnica de València
P. Koehn, University of Edinburgh

November 27, 2012

Index

Work Package 2: Contents

- Goal: Render interaction process flexible and efficient
- Highly active during the first year
- Basic research involving:
  - Novel prediction strategies (task 2.1)
  - Multi-modality (task 2.2)
  - Use of confidence measures (task 2.3)
  - Introduction of syntax-based translation models (task 2.4)
  - Analysis of new SMT models for prediction (task 2.5)
- First year: activity on tasks 2.1, 2.2, 2.4 and 2.5

Introduction

- Despite important advances in SMT, human supervision required for ensuring quality
- Interactive translation prediction (ITP) paradigm:
  - Combine MT systems with human expert knowledge
    - MT system provides efficiency
    - Human expert provides quality
  - User feedback in form of corrections in the translation
  - ITP provides best completion for a partially validated sentence
  - Assumption: translation process is left-to-right

    $\boldsymbol{x}$: sentence in source language

    $\boldsymbol{y}$: sentence in target language

    $\boldsymbol{p}$: validated prefix

    k : current word typed (correction)

    $\boldsymbol{s}$ suffix generated (prediction)

    $(\boldsymbol{p}k\boldsymbol{s}) = \boldsymbol{y}$

ITP Example

SOURCE (x)**:** Para imprimir una lista de fuentes postscript:
REFERENCE (y)**:** To print a list of postscript fonts:

ITER-0	(p)	
ITER-1	$(\hat{s})$	*To print a postscript font list:*
	(p)	To print a
	(k)	list
FINAL	$(\hat{s})$	*of postscript fonts:*
	(k)	$(\#)$
	$(p \equiv y)$	To print a list of postscript fonts:

with $p \equiv$ validated prefix, $k \equiv$ current word typed, $s \equiv$ suffix generated, $(pks) = y$

⇒ Sentence corrected with just one word-stroke (vs. four in post-edition)
⇒ Interaction also possible at the character level

Introduction

- Fundamental equation of MT: $\hat{\boldsymbol{y}} = \text{argmax}_{\boldsymbol{y}}\, p(\boldsymbol{y} \mid \boldsymbol{x})$
  with $\boldsymbol{x}$ the input sentence and $\boldsymbol{y}$ the output sentence

- Leverage fully-fledged SMT systems for ITP:

$$\hat{\boldsymbol{s}} = \underset{\boldsymbol{s}}{\text{argmax}}\, p(\boldsymbol{s} \mid \boldsymbol{x}, \boldsymbol{p}, k)$$

  with $\boldsymbol{p} \equiv$ validated prefix, $k \equiv$ current word typed, $\boldsymbol{s} \equiv$ suffix generated, $(\boldsymbol{p}k\boldsymbol{s}) = \boldsymbol{y}$

- Main problems:
  - Prediction criterion needs to be accurate
  - System cannot explain a prefix given by the user
  - Underlying translation model is transparent, but extremely important
  - HCI must be comfortable and efficient
  - Errors need to be spotted easily and rapidly

Task 2.1

Search and Machine Learning Criteria

Task 2.1: Search and machine learning criteria

- Schedule: months 1–18
- Basic research attempting to improve prediction as such
- Focus: prediction as a machine learning problem
- Two different sub-tasks:
  - Re-consider the prediction problem from a theoretical point of view
  - Research novel error recovery strategies and decision rules

Optimum Decision Rule for ITP: Motivation

- Classical suffix search (as in auto-completion):

$$\hat{\boldsymbol{s}} = \underset{\boldsymbol{s}}{\operatorname{argmax}}\, p(\boldsymbol{s} \mid \boldsymbol{x}, \boldsymbol{p}, k)$$

  – Extension of the classical decision rule for MT
  – Minimises suffix errors (maximum-a-posteriori approach)

- Research questions:
  – Is this rule optimal to minimise interactions?
  – Can we do better?
  – How much?
  – At what cost?

Optimum Decision Rule: Findings

- Optimal decision rule for ITP:

$$\hat{\boldsymbol{s}} = (\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_N) \qquad \text{for some } N$$

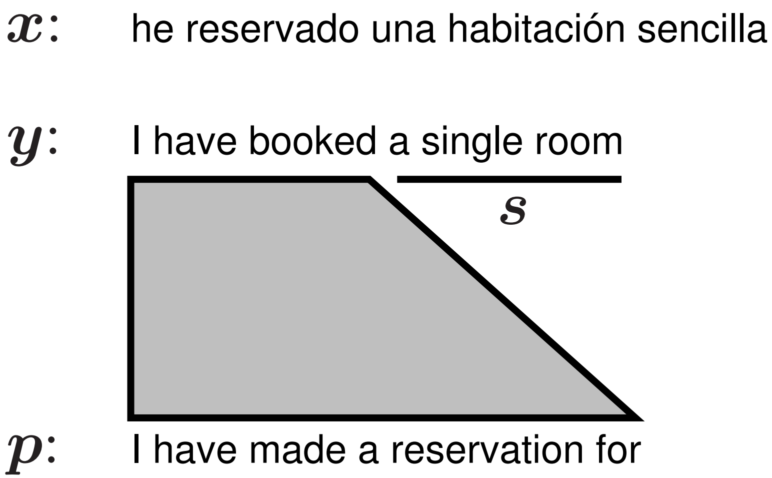
$$\hat{s}_i = \operatorname*{argmax}_{s_i} \sum_{s'} p(\hat{s}_i, \boldsymbol{s}' \mid \boldsymbol{x}, \boldsymbol{p}, k, \hat{s}_1, \hat{s}_2, \ldots, \hat{s}_{i-1})$$

- Findings
  - Suffix is generated by appending one word at a time
  - Efficient algorithm using wordgraphs
  - MAP is a maximum approximation to the sum in optimal ITP
  - Both obtain the same result if maximum dominates

- Results
  - Small but promising improvements for ITP
  - Significant differences in ASR

ITP Based on Stochastic Error-Correction Models

- Regular ITP systems use wordgraphs to generate the required suffixes
- We propose to include error-correction models in the ITP formulation
- The generation of the suffix y can be seen as a two-stage process:
  1. Translate the source sentence x, obtaining y
  2. Align the prefix p with a prefix of y (s is the unaligned portion of y)



Evaluation: ITP with Stochastic Error-Correction Models

- Here we show results obtained with the following ITP systems:
  - ITP system based on error-correction models (ECM)
  - State-of-the-art ITP system using wordgraphs (WG)
  - State-of-the-art ITP system without wordgraphs (NOWG)

- **KSMR** results with the **Xerox** and the **EU** corpora (English-French language pair):

Corpus	**NOWG**	**WG**	**ECM**
En–Fr (**Xerox**)	35.8±1.3	40.4±1.4	37.5±1.2
Fr–En (**EU**)	21.5±1.0	28.6±1.2	23.2±1.0

- ECM was better than other wordgraph-based state-of-the-art systems

- NOWG obtained the best results but ECM was one order of magnitude faster

Task 2.4

Prediction from Parse Forest

Task 2.4: Prediction from Parse Forest

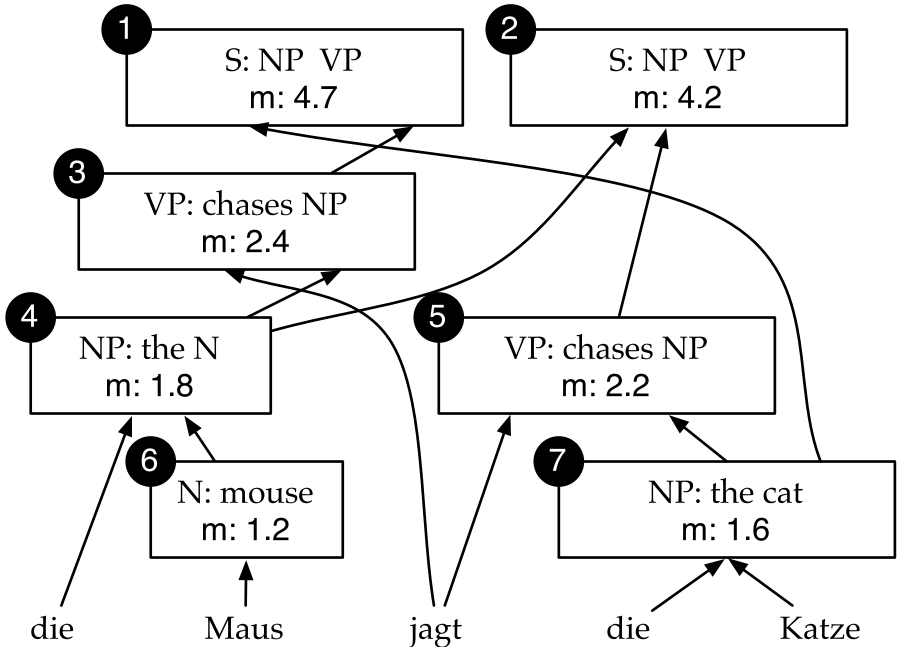
- Schedule: months 6–18
- Enable syntax-based models for ITP
- Focus: Develop prediction algorithms for parse forests
- Syntax-based models have rapidly reached the state-of-the-art
- Syntax-based models naturally evolves away from sequential translation schemes
- Two different sub-tasks:
  - Parse forests obtained from ITGs
  - Parse forests obtained from SCFGs

Prediction from ITG Parse Forest

- ITG-based decoding: CKY traverse
- Syntactic information added
- Translation Parse Forest:
  - Compact representation of context-free translation obtained in decoding
- Completion from human supervised span (not necessarily prefix)
  - Search the node N that represents the span
  - Bottom-up traverse of the forest from N
- Search can also be performed a set of starting nodes
- Pruning of the tree for the next human interaction
- Results in SMT with Zh–En BTEC corpus: 43.0 BLEU (vs. 41.1 with Moses)

Prediction from SCFG Parse Forest

- Parse forest generated by SCFG with target-side syntax

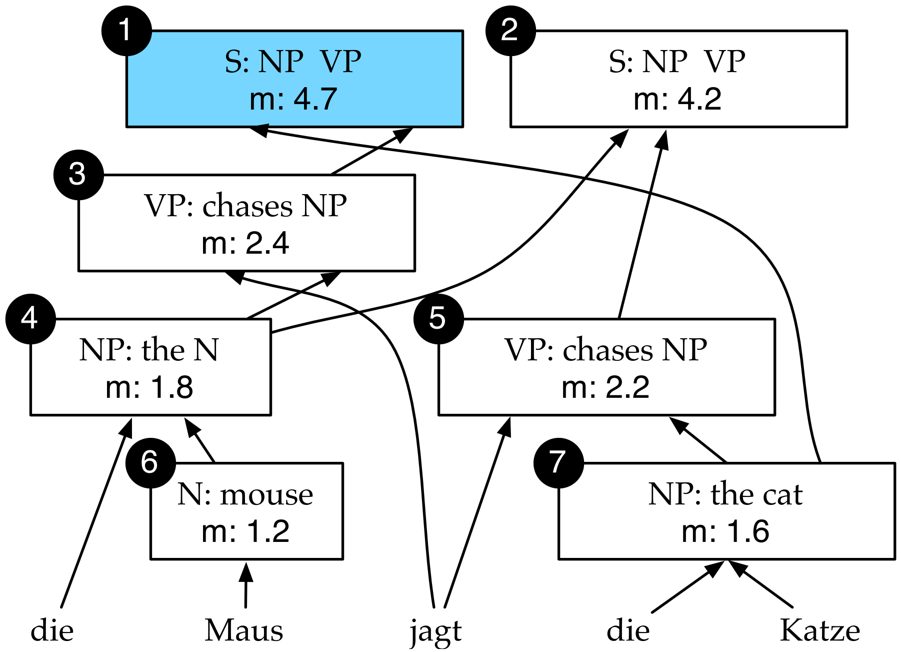


- Task: Optimal completion for the small cat chases

- Main idea: Top down search of hypergraph with dynamic programming

Prediction from SCFG Parse Forest

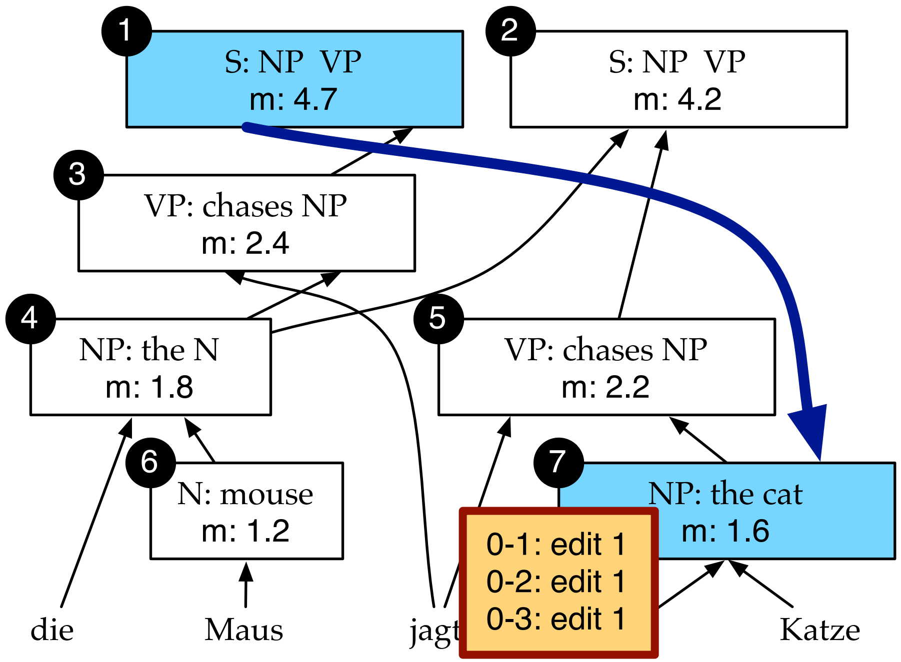
- Start search at top (best full derivation), max edit 1
- No terminals → follow hyper edge



- Task: Optimal completion for the small cat chases

Prediction from SCFG Parse Forest

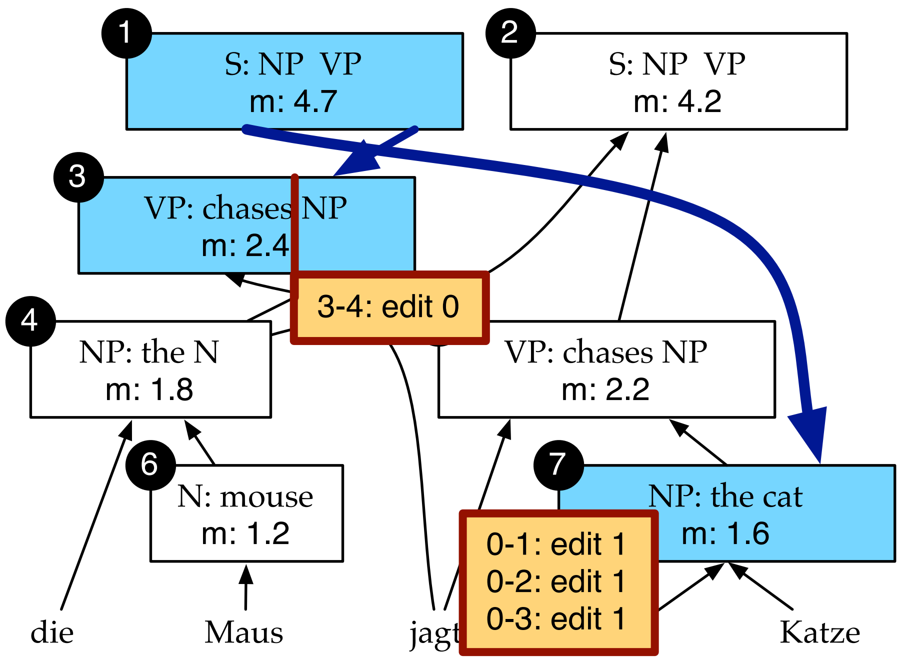
- Non-terminal node, reached with 0 user words consumed
- Compute optimal matches to 1, 2, and 3-words of user prefix (max edit 1)



- Task: Optimal completion for the small cat chases

Prediction from SCFG Parse Forest

- Back to top node, follow second hyperedge (already 1 edit, now 0 edits allowed)
- Compute only allowed edit, matches node partially

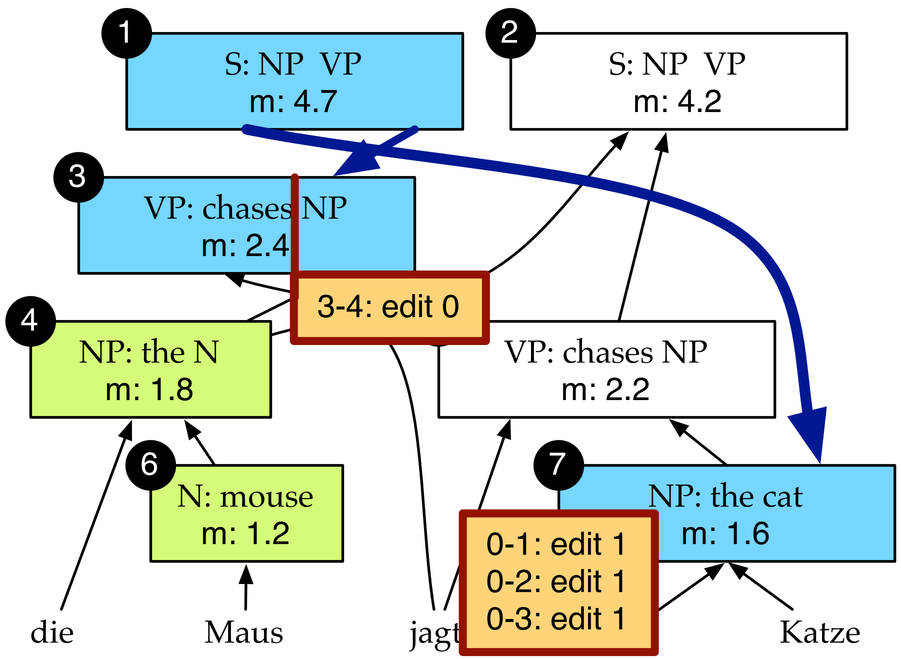


- Task: Optimal completion for the small cat chases

Prediction from SCFG Parse Forest

- After all user words consumed

⇒ read out remaining words of derivation: the mouse.



- Task: Optimal completion for the small cat chases

Prediction from SCFG Parse Forest

- Incrementally increase maximum edits from 0, recompute
- Follow full derivations top down
- Reach each node with maximum edits left
- Compute optimal edits to match substring of user input
- Cache results for each node

- Currently working on refinements and experiments

Task 2.5

New SMT Models for ITP

Task 2.5: New SMT Models for ITP

- Schedule: months 1–12
- Basic research concerning the core of ITP: SMT models
- Focus: Develop novel SMT models for their application to ITP
- Recent developments: HMMs and finite-state models could provide improvements
- Study their applicability within the ITP framework

PB Hidden Semi-Markov Models

- Advantage: strong theoretical foundation
- Disadvantage: high computational cost in training $\mathcal{O}(I^2 \cdot J^2)$
- Results on Europarl-v3

System	Distance-based reordering En $\rightarrow$ Sp	
	TER	BLEU
Moses	55.0 ± 0.8	29.9 ± 0.9
log-PBHSMM	55.6 ± 0.9	29.3 ± 0.9
Moses+log-PBHSMM	54.9 ± 0.9	29.9 ± 0.8

- Competitive with Moses
- Training time too expensive
  $\Rightarrow$ Not implemented in ITP

A Finite-State Approach to Phrase-Based SMT

- Finite-state approach
- Seven monotone models in log-linear combination
- All models implemented by means of WFSTs
  - Translation WFST $\mathcal{T}$ combined on the fly with LM WFST $\mathcal{L}$, i.e. $\mathcal{T} \circ \mathcal{L}$.

Experiments on an English–French TED translation task

System	BLEU
PBFSMT	22.2
Monotone Moses	22.0
Non-monotone Moses	23.2

- Comparable to monotone Moses bu not to non-monotone Moses
  $\Rightarrow$ Not implemented in ITP

Task 2.2

Multi-Modality in Translation Prediction

Task 2.2: Multi-Modality in Translation Prediction

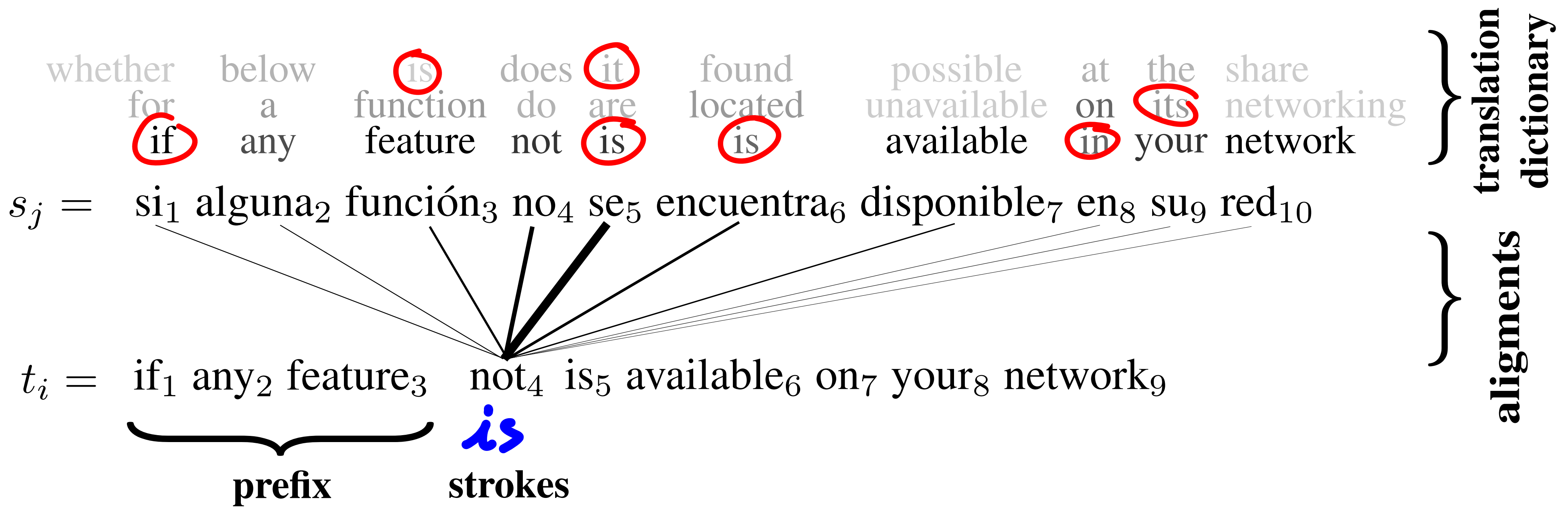
- Schedule: months 6–24
- Introduce new technologies into the interaction framework
- Focus: design a comfortable and efficient user interface
- Interaction with mouse and keyboard is intuitive, but often simple and inefficient
- Introduce other intuitive HCI devices: e-pen or touch-screens

Study of E-pen Gestures for ITP

deletion	if_1 any_2 feature_3 not_4 is_5 available_6 on_7 your_8 network_9
insertion	if_1 any_2 feature_3 not_4 is_5 available_6 on_7 your_8 network_9
substitution	if_1 any_2 feature_3 not_4 is_5 available_6 on_7 your_8 network_9
shift	if_1 any_2 feature_3 not_4 is_5 available_6 on_7 your_8 network_9
transposition	if_1 any_2 feature_3 not_4 is_5 available_6 on_7 your_8 network_9

TER

Leveraging MT Context in Online HTR



- Leverage information from:
  - Target prefix
  - Word-based translation models

Online HTR Results

- results for Xerox corpus
- edit rate: $\frac{\text{n. operations}}{\text{n. words reference}}$
- ITP ~ 30 more handwriting than PE
- ITP avoids deletions

System	es (ER %)	en (ER %)
post-editing	24.4	21.3
ITP	22.8	21.1

HTR : baseline

4**PREF** : 4-gram prefix

M2 : word model 2
(dictionary + alignments)

System	es (CER %)	en (CER %)
HTR	11.1	9.9
4**PREF**	9.9	9.5
M2	**8.6**	7.7
M2+4**PREF**	9.0	7.5

Conclusions

- Task 2.1:
  - Optimal decision rule can be implemented efficiently by means of word graphs
    $\Rightarrow$ Limited improvements in ITP, but theoretically sound
  - Stochastic error-correcting proves to outperform other systems
- Task 2.4: Work in progress for implementing ITP by means of ITGs and SCFGs
  - ITG toolkit extended to produce parse forests
  - On-going experiments regarding prediction from SCFG parse forest
- Task 2.5: PBHSMM and PBFSMT analysed for applicability to ITP
  - Not implemented for ITP due to different constraints
- Task 2.2: Use e-pen for interacting with PE or ITP system
  - Correction by on-screen writing
  - Correction by means of e-pen gestures
- Task 2.3: Just started

Future Work

- Conduct experiments with standard CASMACAT corpora
- Implement weaker feedback into the CASMACAT workbench
- Task 2.1:
  - Transform optimum decision rule into greedy version
  - Extend stochastic error-correcting by means of IBM alignments
- Task 2.4: Implement prediction from ITG and SCFG parse forests
- Task 2.5: task completed
- Task 2.2:
  - Allow multi-word corrections
  - Implement within CASMACAT workbench

Publications

- V. Alabau and F. Casacuberta. Study of Electronic Pen Commands for Interactive-Predictive Machine Translation. Proc. of the Intl. Workshop on Expertise in Translation and Post-editing Research and Application, 2012.
- J. González-Rubio, A. Sanchis and F. Casacuberta. PRHLT Submission to the WMT12 Quality Estimation Task. Proc. of the Seventh Workshop on SMT, North American Chapter of the Association for Computational Linguistics (NAACL), pp. 104-108, Association for Computational Linguistics, 2012.
- V. Alabau, A. Sanchis and F. Casacuberta. On the Optimal Decision Rule for Sequential Interactive Structured Prediction. Pattern Recognition Letters, 33:6, pp. 2226-2231, 2012.
- G. Sanchis-Trilles. Building task-oriented machine translation systems. PhD Thesis, Universitat Politècnica de València, 2012.
- J. González. A finite-state approach to phrase-based statistical machine translation. Proc. of the 10th edition of the Intl. Workshop on Finite State Methods and Natural Language Processing, pp. 95-103, 2012.