

System Architecture

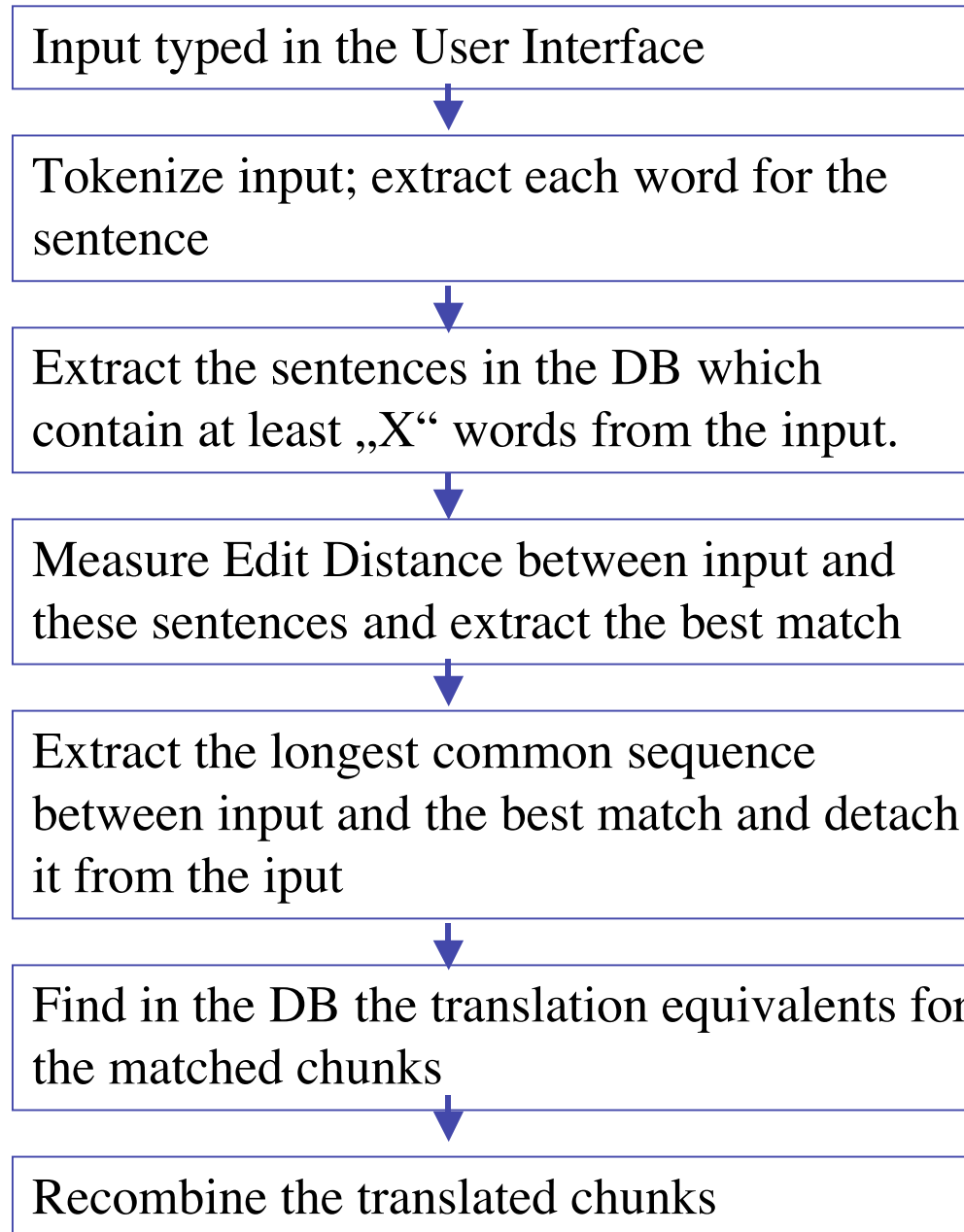
Cristina Vertan

vertan@informatik.uni-hamburg.de

Contents

- System Architecture
- Workflow
- Resources
- Evaluation

Workflow first version 23.11.2005



Repeat the operation with rest of the input

Tasks first version

-User Interface Group-

- Fix with the other groups the communication interfaces (number and type of arguments, access methods) between all modules
- Find a tokenizer for German, one for English and integrate it with the system
- Insert the links between the DE and EN lexicon
- Provide methods for accessing the lexicon

Tasks first version -Database Group-

- Organise the DB for fast search
- Provide methods for accessing the DB, and retrieval of candidate solutions for matching
- Implement a method for finding the translation equivalents through simple word-by-word look-up in the dictionary

Tasks first version

-Matching group-

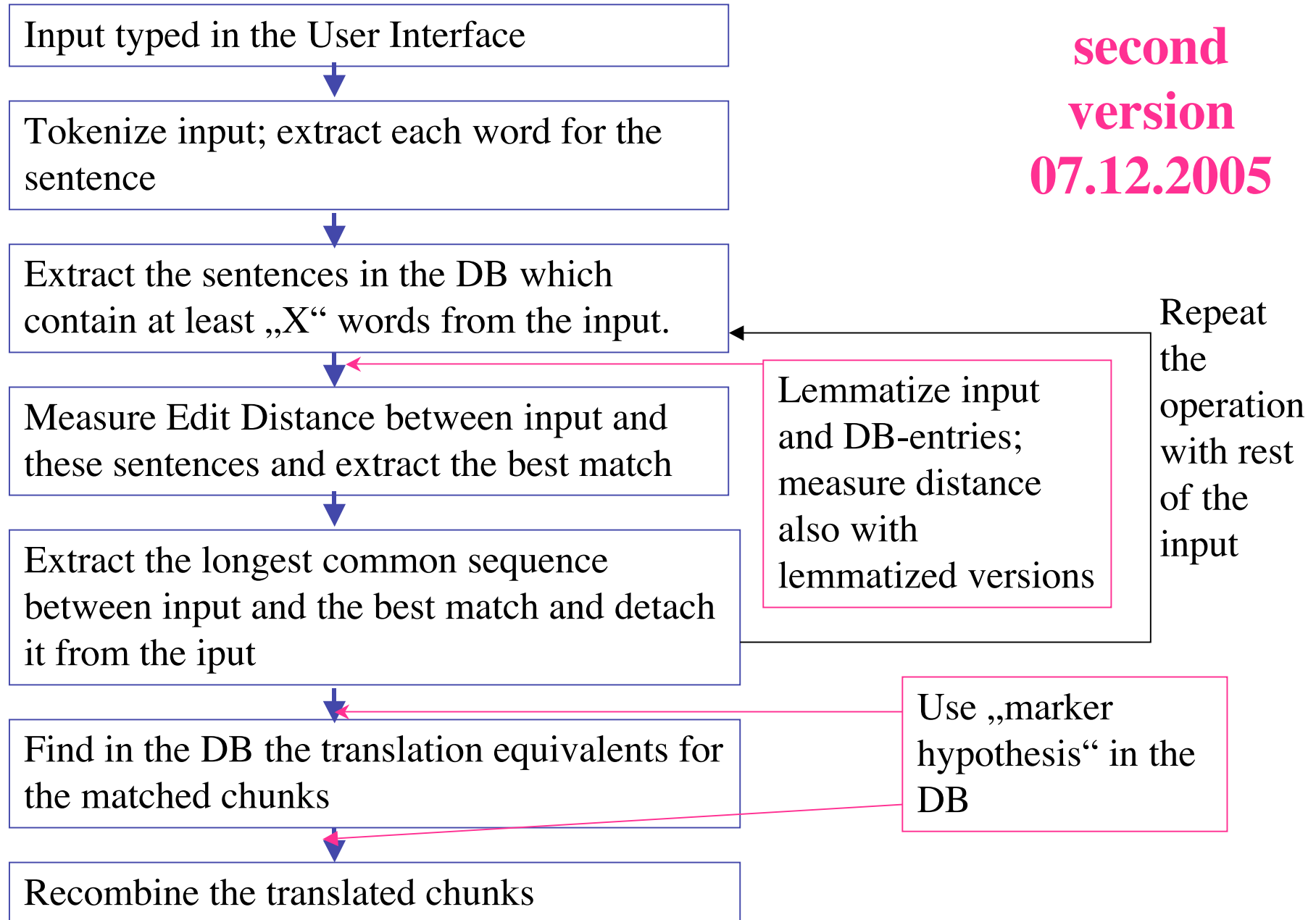
- Implement edit distance
- Fix a treashold for the edit distance
- Implement longest common sequence

Tasks first version

-Recombination Group-

- Define simple rules for recombination of translation segments (e.g. look-up at POS in Lexikon and definition of simplerules like Adj N --> N Adj)
- Implement the recombination procedure
- Provide results to the User Interface

**Workflow
second
version
07.12.2005**



Tasks second version

-User Interface Group-

- Implement BLEU method

Tasks second version

-Database Group-

- Find a method to lemmatize the input and the DB-entries in SL (use/add information in lexicon or use a lemmatizer)
- Insert lemmatized versions of the entries in the DB
- Use marker Hypothesis and mark the DB

Tasks second version

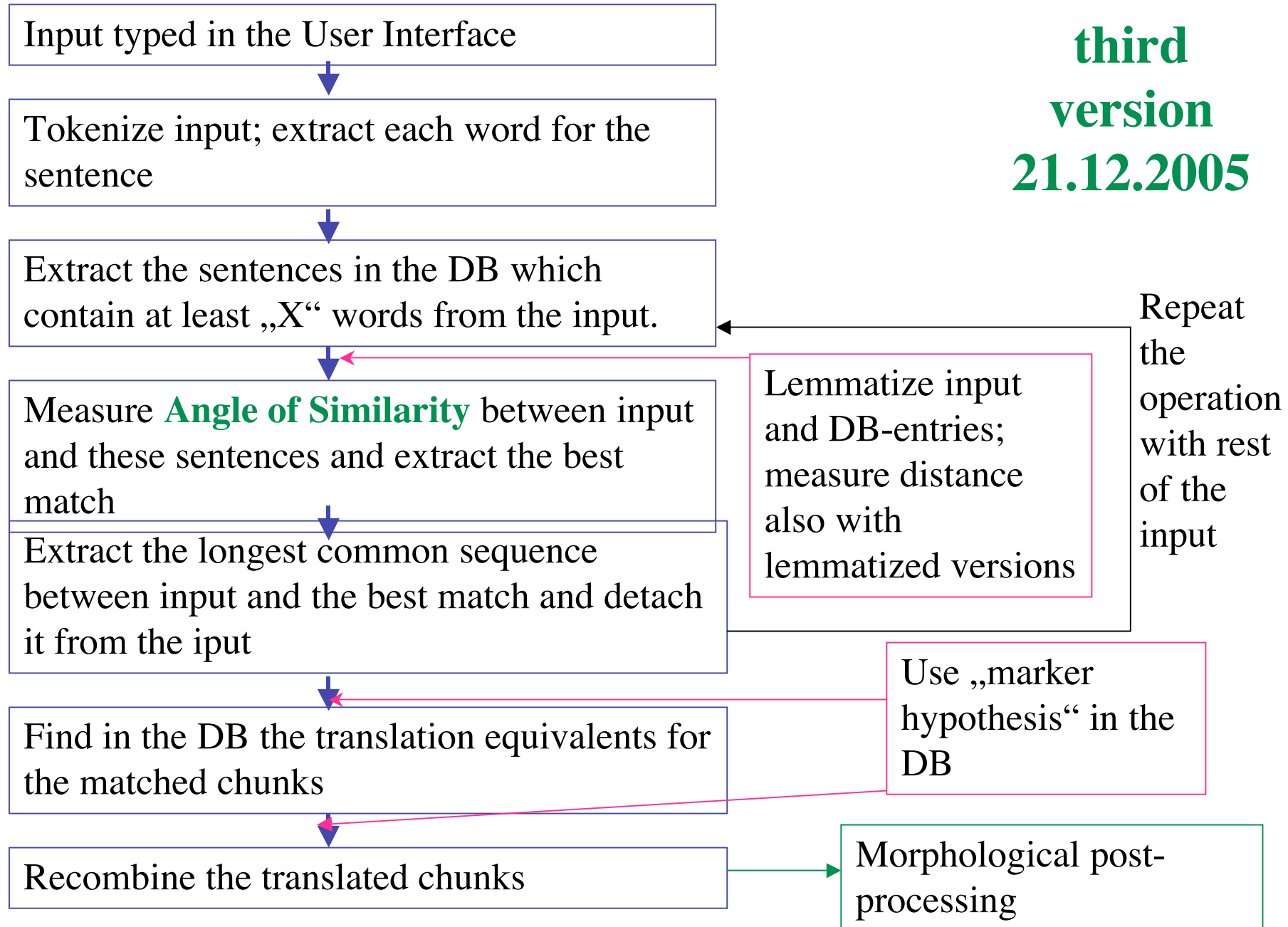
-Matching group-

- Extend your matching algorithm so that not only the strings but also the lemmatized versions are compared.
- Fix some criteria to accept/reject the candidate DB-entries

Tasks second version -Recombination Group-

- Improve the recombination method through use the marker hypothesis
- Replace untranslated chunks through their word-by word translation

Workflow third version 21.12.2005



Tasks thirdversion

-User Interface Group-

- Adapt the Evaluation Tool in order to read the input and the provided translation
- Provide the reference translations for the test sentences

Tasks third version

-Database Group-

- Extend the Database
- Extend the DE and EN lexikon as well as the links

Tasks third version

-Matching group-

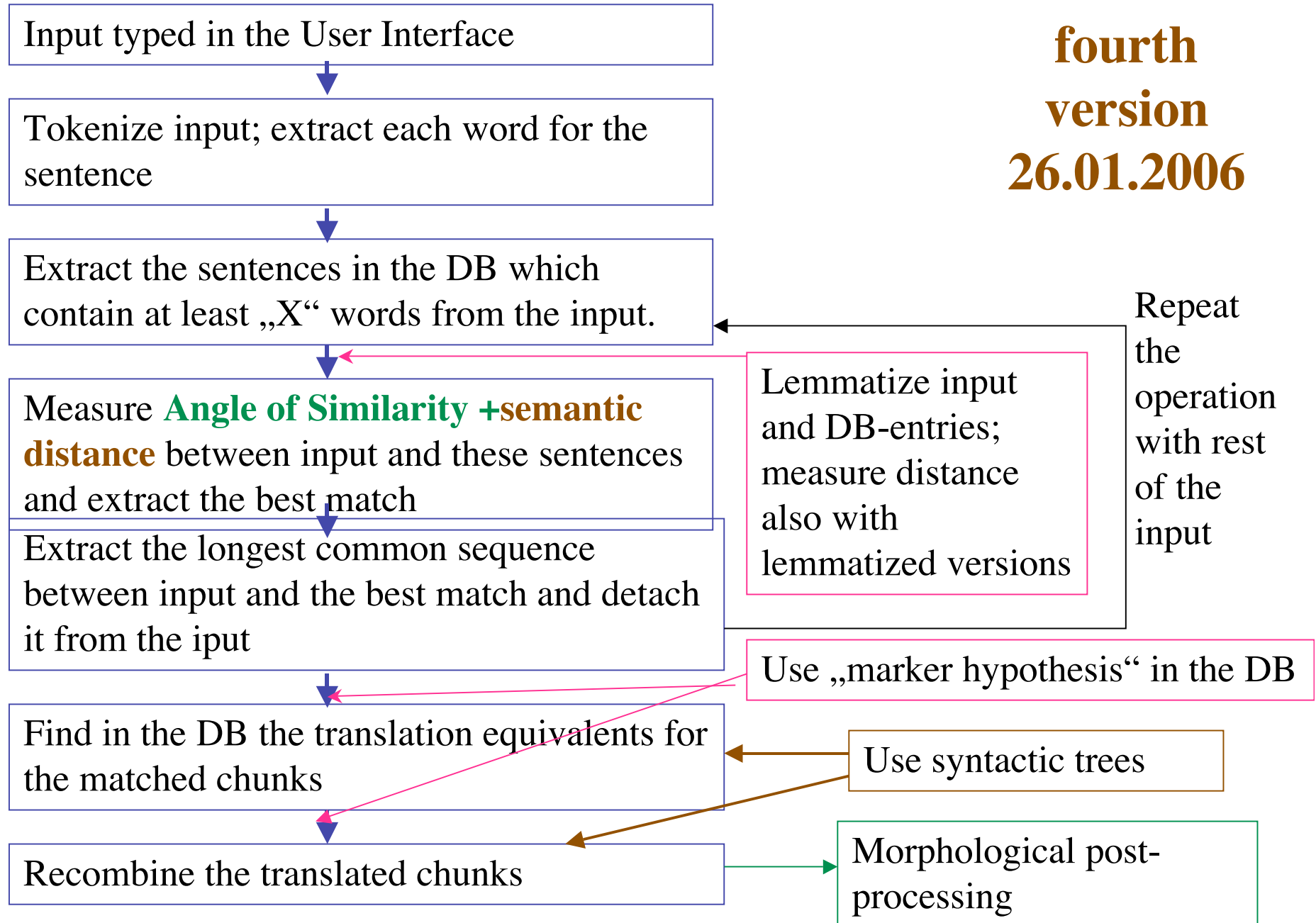
- Implement angle of similarity

Tasks third version

-Recombination Group-

- Implement/ or integrate some morphological generation for the output

Workflow fourth version 26.01.2006



Tasks fourth version

-User Interface Group-

- Integrate BLEU in the evaluation tool

Tasks fourth version

-Database Group-

- Integrate some syntactic trees for sentences in the DB

Tasks fourth version

-Matching group-

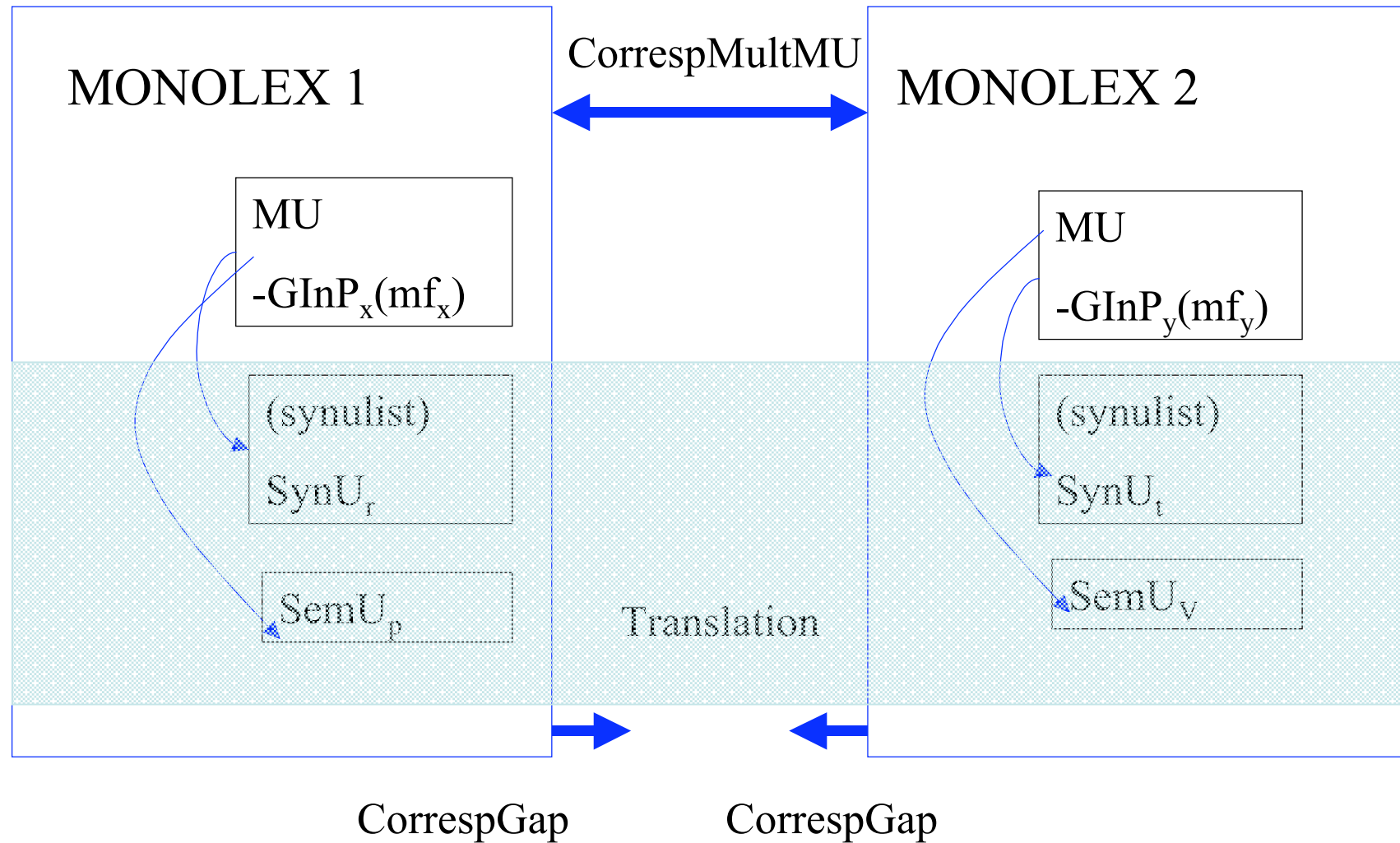
- Design a semantic network for technical nouns
- Implement semantic distance and integrate with the matching algorithm

Tasks fourth version

-Recombination Group-

- Use the associated trees for the retrieved chunks for the recombination

Lexicon Architecture



GET

File Settings View Statistics Documents

Wir treffen uns vor der Pizzeria Lorenzo. Ein italienisches Restaurant.

we meet in front of the seats and that way an Italian restaurant

Translation Mismatch
 No Yes

Translation Soundness
 Machine Human

Translation Quality
 Good Intermediate Bad

Yes No

Input Yes No

Yes No

Syntactically Correct Yes No

Semantically Correct Yes No

Possible Misunderstandings Yes No

Output

How long is the drive to Hanover?

wie Lange fahren wir nach Hannover

Next Turn

Information Elements

Lost Information Elements

Essential Information Elements

Translated Information Elements

Added Information Elements

Turn Number

File text2.eval
 Turns loaded 27

Evaluation Tool

Recognized input

Original input: Circa zwei Stunden. Wir sind dann um neun Uhr in Hannover.

Recognized input: zirka zwei Stunden Wiedersehen dann um neun Uhr in Hannover

Buttons: Compare, Compare all German turns, Compare all English turns, Dismiss

1 Cased based translation: two hours i will be in Hanover at nine o'clock .

Statistical translation: about two hours then at nine o'clock , in Hanover .

Dialog based translation: it lasts for two hours at nine o'clock in Hanover

Deep analysis

Buttons: Update, Dismiss

Dialog act

Original: Circa zwei Stunden. Wir sind dann um neun Uhr in Hannover.

Buttons: greet, bye, introduce, politeness formula, thank, deliberate, backchannel, **init**, defer, close, commit, offer, request suggest, request clarify, request comment, request commit, suggest, inform digress, inform exclude, inform clarify, inform give reason, accept, confirm

Statistics

Criteria	Percentage	Value
Good	40.74074%	(11.0/27.0)
Intermediate	18.518518%	(5.0/27.0)
Bad	40.74074%	(11.0/27.0)
Syntactically correct	44.444447%	(12.0/27.0)
Semantically correct	51.851852%	(14.0/27.0)
No possible misunderstandings	55.555557%	(15.0/27.0)
Translation mismatch	59.25926%	(16.0/27.0)
No translation mismatch	37.037037%	(10.0/27.0)
Full dialog act preservation	0.0%	(0.0/27.0)

Buttons: update, Dismiss

GET

Statistics Documents

text1.eval
 text2.eval

Mir sind dann über.

Translation Mismatch: No Yes

Translation Soundness: Machine Human

Translation Quality: Good Intermediate Bad

it lasts for two hours at nine o'clock in Hannover

Syntactically Correct: Yes No

Semantically Correct: Yes No

Possible Misunderstandings: Yes No

Output

hotel have a ent?

Next Turn

sollen wir das Hotel mit Bar nehmen

Lost Information Elements: - 2 +

Translated Information Elements: - 3 +

Added Information Elements: - 0 +

Turn Number: < 16 >

File text2.eval
Turns loaded 27