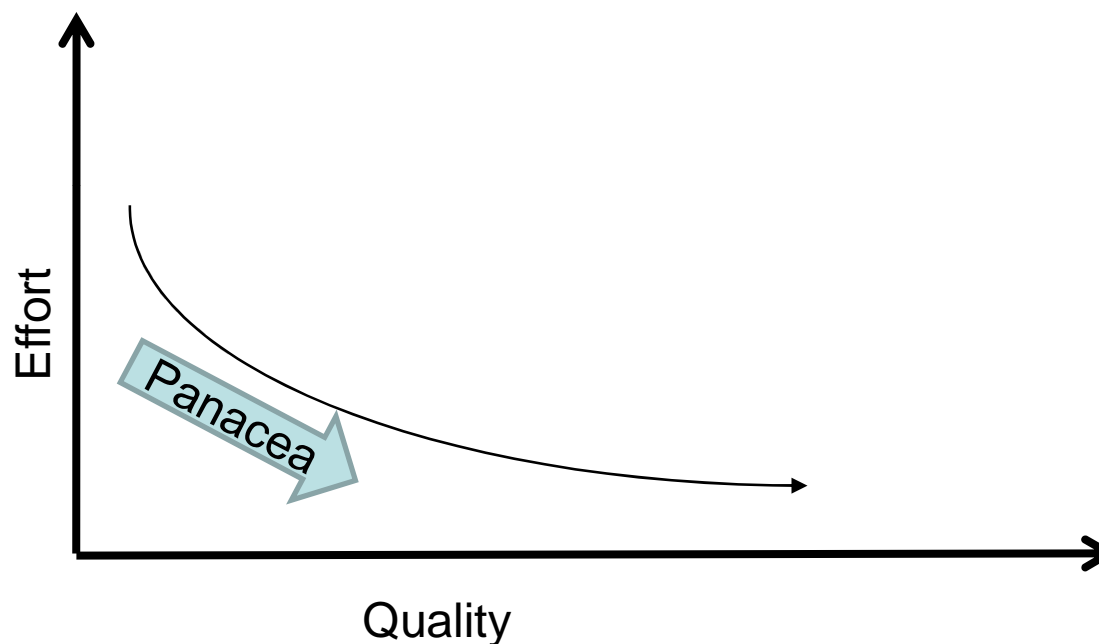




# PANACEA WP8: Evaluation in industrial environments

Barcelona, 2010-01-21  
UPF, DCU, ELDA, LG  
G. Thurmair, Linguattec

- Relationship between investment (effort) and return (well, ... quality)



PANACEA target:

Productivity gain = *better* quality with *less* effort



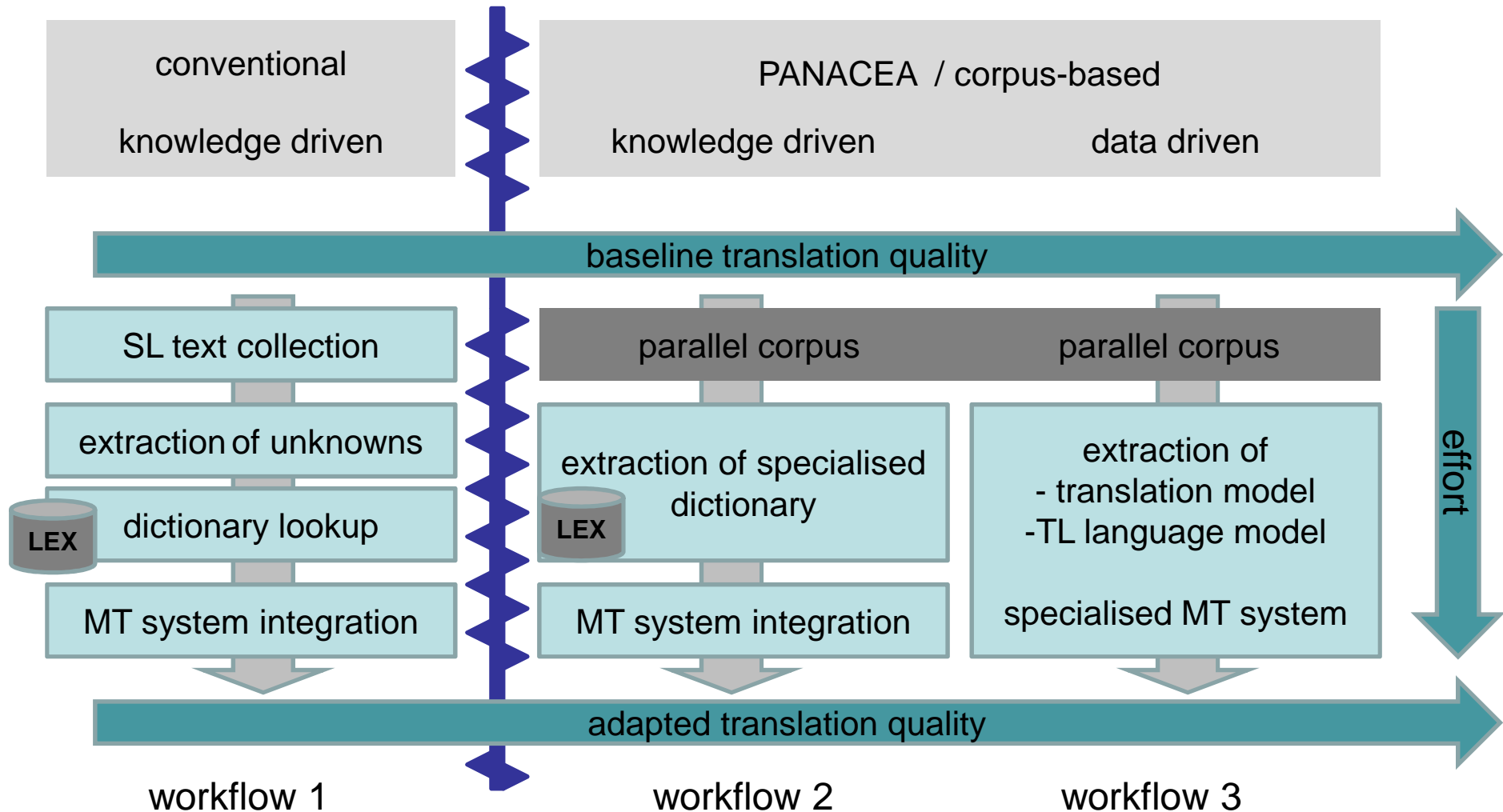
# Tasks of WP 8



- Evaluation of *quality* and *productivity* of the PANACEA tools
- Subtasks
  - define industrial setup for LR production
    - user / application requirements T6
  - tool-oriented evaluation
    - quality, usability of tools in such applications T36
  - task-oriented evaluation
    - typical industrial workflow T36

# Industrial Setup 1

- Selection of industrial applications:
    - Machine Translation (multilingual)
      - adaptation of MT system to a specific domain
        - focus on: translation (bilingual terminology)
    - Alerting system (monolingual)
      - specification of LRs for alerting system
        - focus on: extraction & search (concepts, proper names)
- => focus on MT in Panacea-WP8





# Machine Translation



- Task
  - adaptation of MT system to new domain
    - domain to be defined; candidates:
      - automotive / medical / assistive tech / bioscience
- Methodology
  - evaluate quality gain against production effort
    - Step 1: create baseline translation
    - Step 2: develop adaptation LRs
      - in conventional workflow
      - PANACEA-based
    - Step 3: measure quality improvement and production effort

# Step 1

- **Collection of test corpus**
  - (no reference translations available ...)
  - several K sentences
    - from the domain
    - out of domain
- **Creation of baseline MT output**
  - for SMT system (MaTrEx fallback: Moses)
  - for RMT system (Linguec Personal Translator)

## Step 2

- Creation of training data
  - monolingual corpora (workflow 1)
  - bilingual corpora (workflow 2 and 3)
- development of three adapted systems
  - conventional
    - unknown word search, dictionary work
  - adaptation of general (RMT) system
    - additional PANACEA special domain dictionary
      - » using PANACEA extraction and dictionary tools
  - new specialised MT system (data driven)
    - DCU MaTrEx (existing technical SMT platform)
      - » using PANACEA aligners, chunkers etc.
- Keep track of efforts needed for the adaptation



# Step 3

- Evaluation
  - compare the 3 outputs for quality
    - output inspection (domain expert)
  - quality measured as
    - accuracy, measured in lexical evaluation
      - How many SL concepts are found in the TL output
        - » concepts = terms, names etc.
    - fluency, measured in dependency-based phrases
      - How many well-formed phrases/chunks are found in TL output
  - comparison
    - quality gain compared to baseline translation
    - relation of quality gain and effort (= required investment)

- Existing automatic metrics are not the best option
  - treat RMT and SMT systems differently (ACL2009)
  - in our task-based setup: no reference translations!
- Instead: Use FEMTI criteria: **fluency, accuracy**
- **Accuracy**: related to **concept integrity** (carry concepts from SL to TL)
  - multilingual WSD, contrastive lexical evaluation  
(Apidianaki 2008,2009, Max et al. 2010)
- **Fluency**: related to **grammaticality** (How grammatical is the TL text?)
  - use parsers to decide this: on phrase / chunk level  
(Owczarzak et al. 2007, related work)
- Approach: Tool-supported human evaluation



# Next Steps



- Definition of user requirements (T6 report)
  - based on the workflow just outlined
  - including
    - selection of domain and corpus resources
    - specification of Step 2 tasks
    - review and better definition of evaluation criteria
      - specification of evaluation tools
    - definition of test scenario