



TQPro: Translation Quality for Professionals

www.tqpro.de

Gr. Thurmair

May 2001



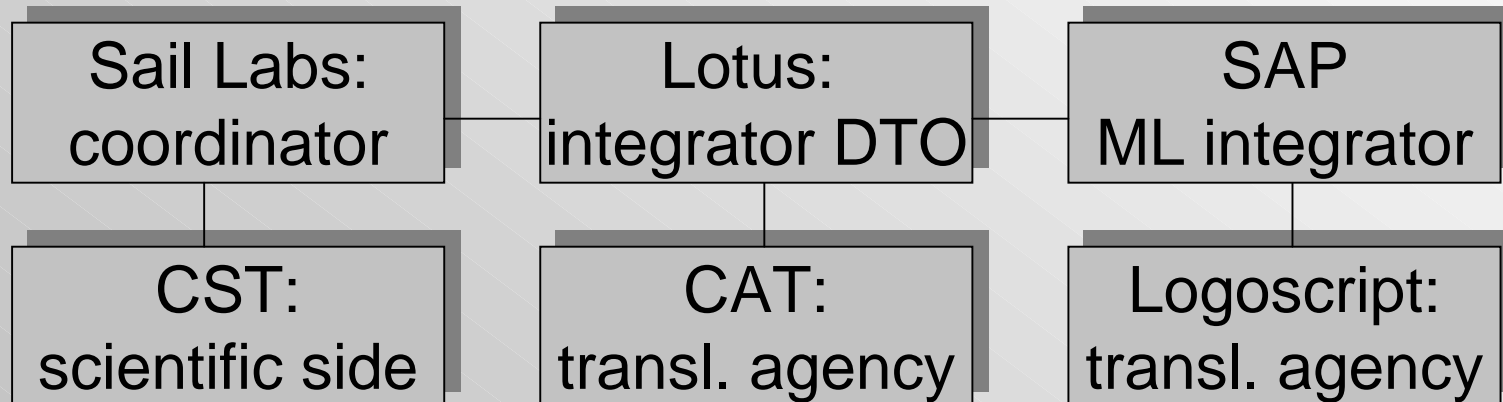
Project Objectives



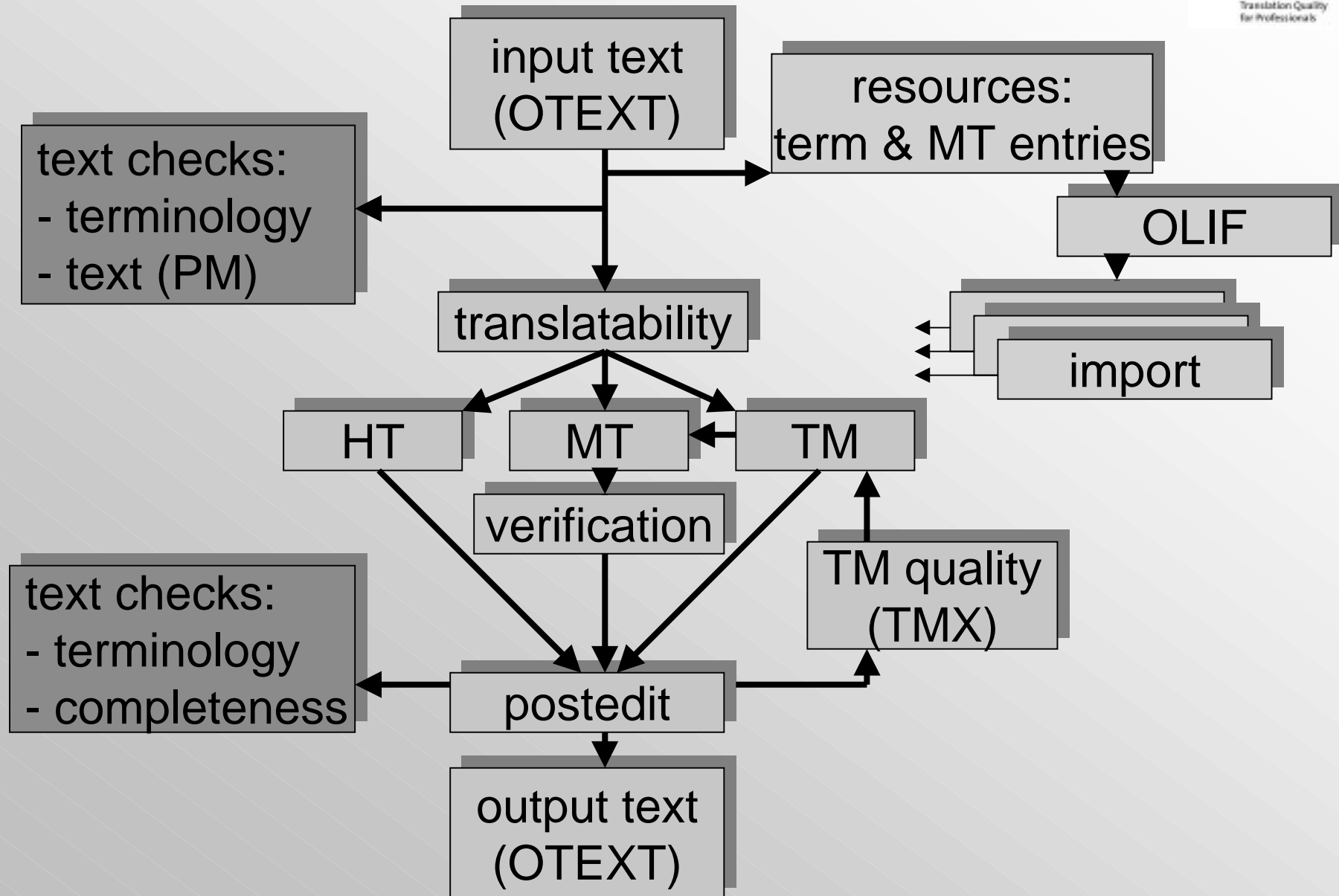
- improve translation **productivity**
 - reduce ramp-up time for resources
 - provide pre- and post-editing functionality
- improve translation **quality**
 - check quality of memories
 - improve machine translation quality

TQPro Managerial Issues

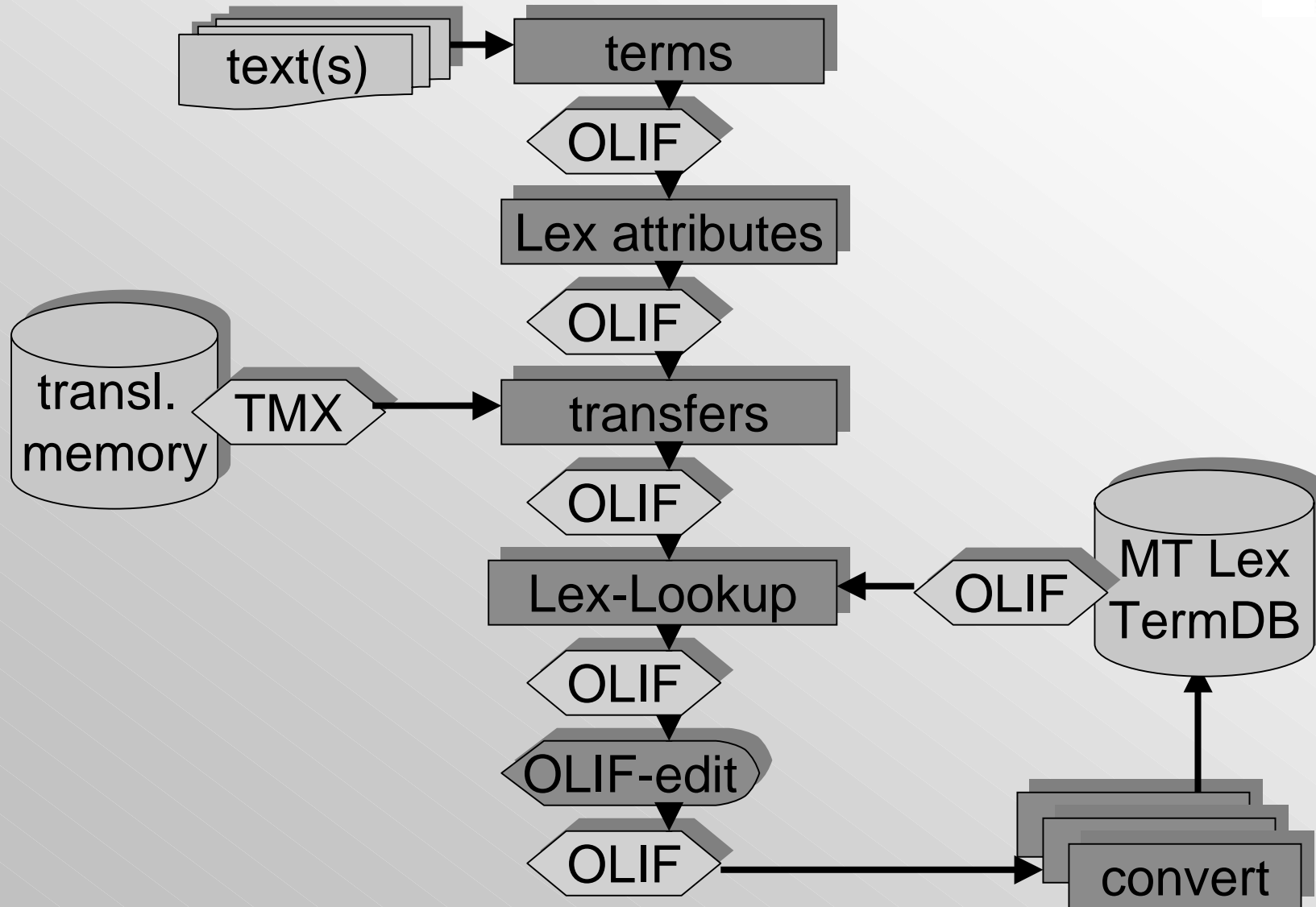
- Duration: 4/00 to 9/02
- Cost. ~ 3 mio, funding 1.6 mio
- Consortium:



TQPro Virtual workflow



Resource creation (WP3)



Preparation Tools (WP 4)



- Is the input text terminology correct?
 - Monolingual Term Verification:
 - compare with lexicon of canonical terms
 - (normalise on text and lexicon side)
- Does the text need preparation?
 - Intelligent Pattern Matching
 - mark up / correct special text portions
(constants, acronyms, measures, filenames, ...)
 - regular expressions and list lookup,
 - GUI for definition of expressions

Preparation Tools (WP 4)



- Can the text be translated by MT?
 - translatability check
 - MT system independent
 - based on shallow parsing and syntactic patterns
 - English, possibly German
 - lexicon-independent (coding effort)

Revision Tools (WP 6)



- Has the intended translation been used?
 - bilingual term verification
 - search source term, check used equivalents
 - TMX format (memory check)
 - flag unknown equivalents
 - (test 1:n options)
- Do text portions need to be changed?
 - intelligent search & replace
 - do morphological analysis on search pattern
 - do proper inflection and agreement changes
 - replace „Fabrik -> Studio“ =>*
 - in der schoenen Fabrik -> in dem schoenen Studio*

Revision Tools (WP 6)



- Is the translation complete?
 - completeness checker
 - are all segments there?
 - have they been translated? (language identifier)
 - is the formatting correct (heading hierarchies etc.)
- Where do I have to postedit first?
 - MT output diagnosis
 - flag critical translation units
 - do this in an MT system independent way:
 - target language model, probabilities of sequences

Translation Tools: Memory

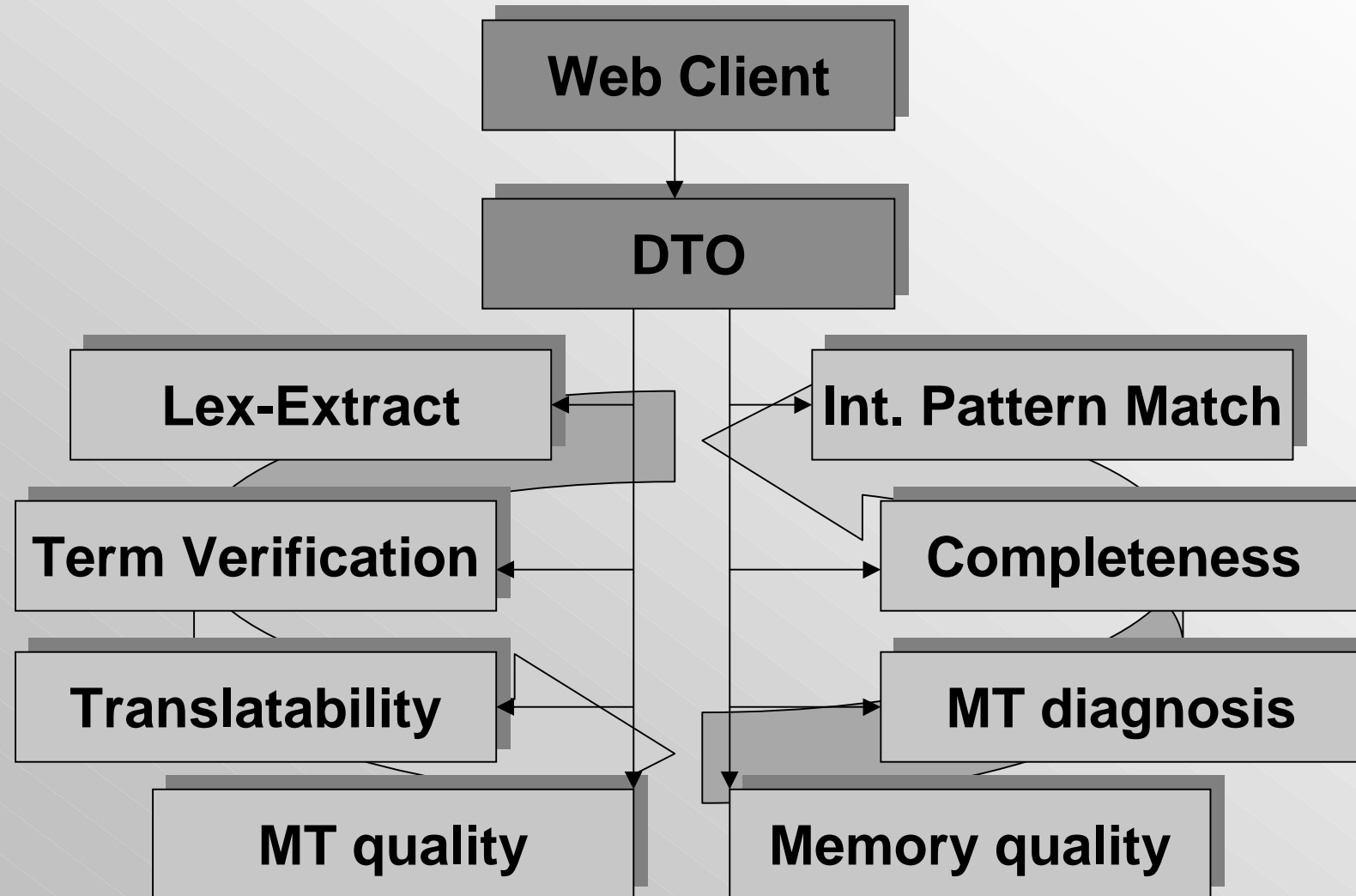


- Improve quality of translation memory:
 - find untranslated / „strange“ segments
 - find improper terminology
 - bilingual term verification:
memory - term base comparison
 - find mis-aligned segments
 - find „similar“ segments
 - based on dependency structure analysis
- run this „memory quality check“
before uploading a finished translation

Translation: MT Quality

- Experiments for MT quality improvements
 - statistical techniques
 - rule frequency in analysis
 - statistical smoothing of generation
 - semantic information
 - SIMPLE semantic categories for disambiguation
 - conceptual semantics in analysis
- realistic size corpora
 - 3 domains, > 3000 sentences, 2 languages
- comparison with existing output quality

Integration: (WP 7)



4 Interfaces and standards

- Text Handling: **OText**
 - markup structure for texts
- Memories: **TMX**
- **MT API**
 - call MT engines in multi-vendor environments
(Lotus DTO, L&H iTranslator)
- Lexicon interchange: **OLIF**
 - exchange of lexical material

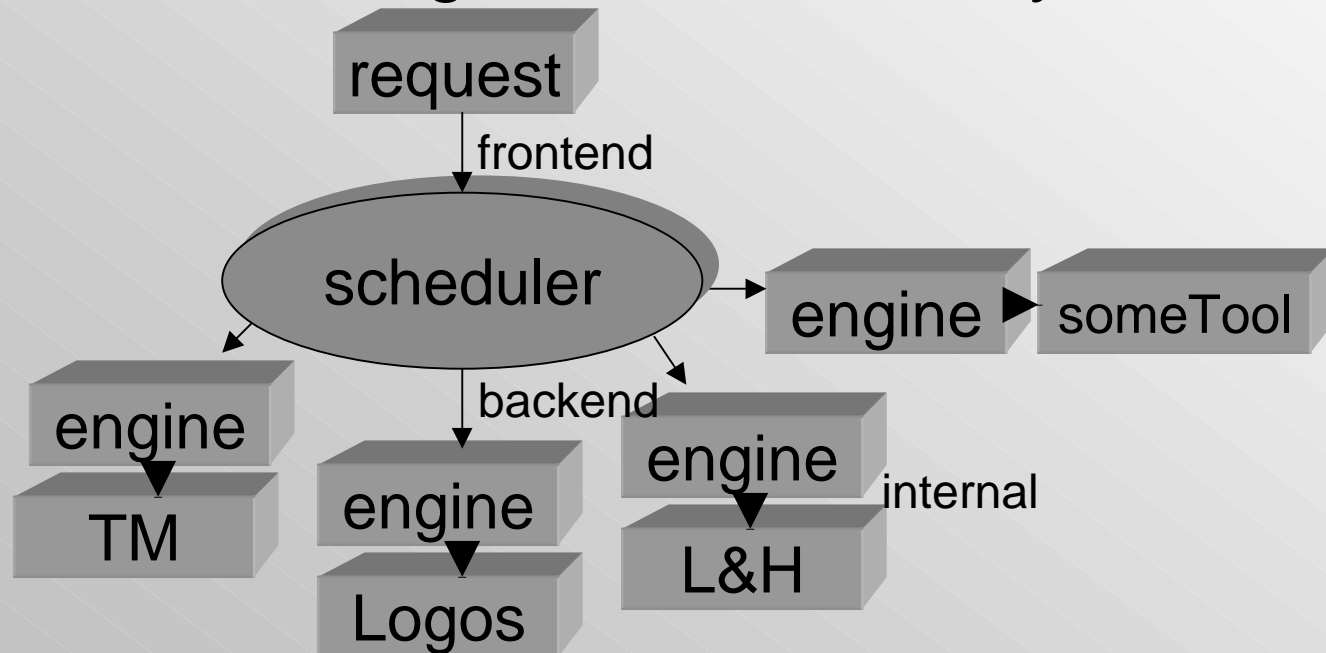
OText: Text Handling



- lean XML format for text handling
 - paragraphs (language, type, flow)
 - text units + variants (language, creator)
 - fonts (like in OpenTag, not TMX!)
 - literals (person, date, email, ...)
 - phrases + variants (index term, term translation)
 - comments, protected, ...
- used by Lotus DTO, Sail
- text format for all TQPro tools
 - (except formatting part of completeness check)

MT APIs

- Three types of APIs
 - „frontend“: client submits a job to scheduler
 - „backend“: scheduler calls an engine
 - „internal“: an engine calls an MT system



OLIF: Lexicon Exchange



- Exchange of lexical resources between
 - different applications, different vendors
SAP, IBM, Systran, Logos, Sail, Xerox, Microsoft, Trados, ...
- Lean exchange, flat structure, XML
 - focus on lexical material, some terminological
- Based on semantic units (concepts)
 - descriptions for
 - monolingual properties (lex, term)
 - transfers (multi-bilingual)
 - crossreferences (ontologies, hierarchies)