

Management, Tasks and Timeline

Linguistic Core MT & TA

Primary Task Breakdown (as proposed)

Task 1

- 📌 Design core linguistic modules and rule formalism, elicitation, contrastive analysis, etc.

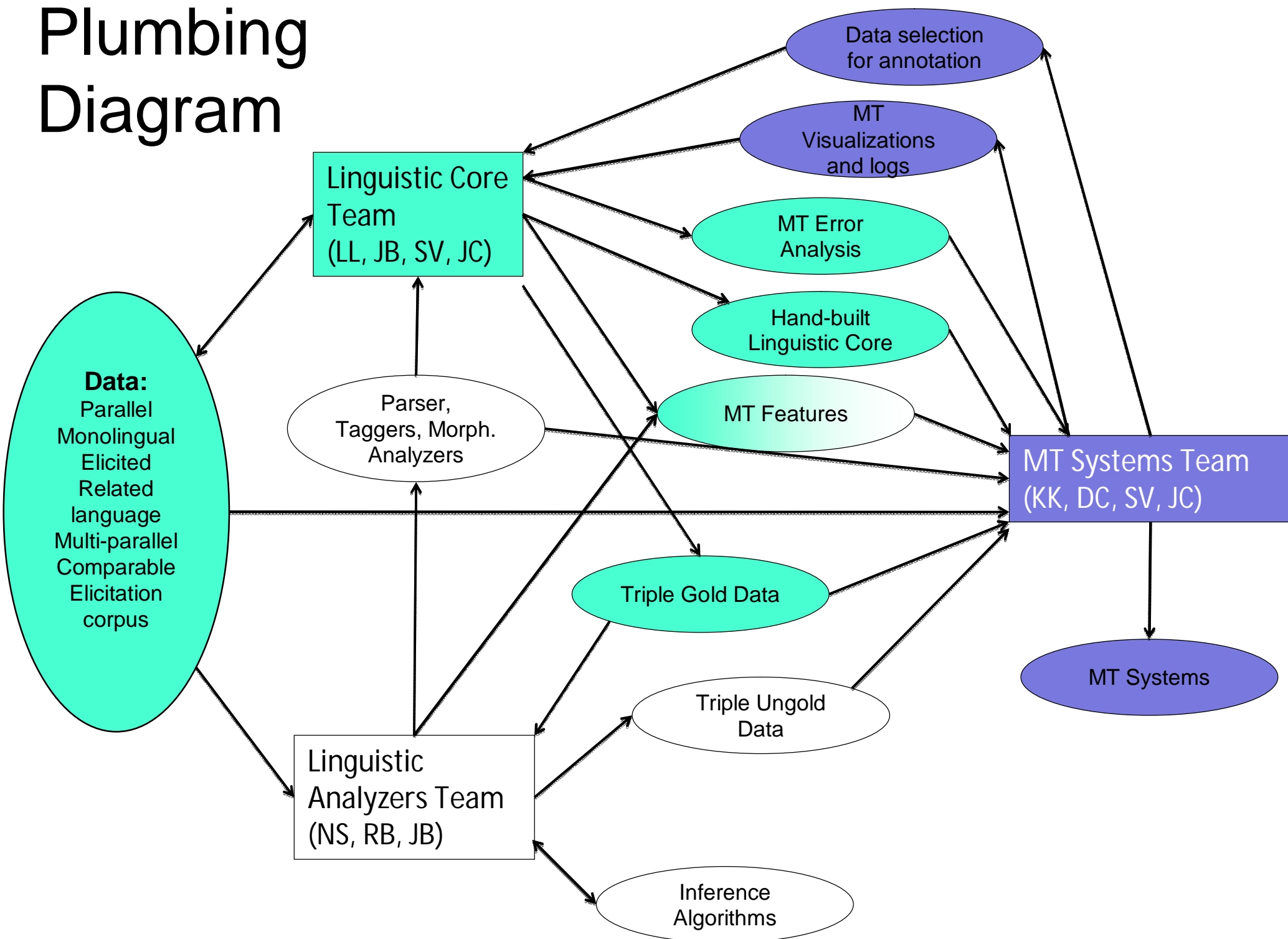
Task 2

- 📌 Methods and algorithms: tree-transduction dependency parsing, learning methods

Task 3

- 📌 MT and TA frameworks, glass-box prototypes, evaluations

Plumbing Diagram



Management and Communication


- Tightly-meshed teams across institutions
 - Bi-weekly coordination conf calls
 - Project wiki, web-pages, code-base, ...
- Explore alternative and multi-faceted methods
 - Contrasted, selected, united, ...
- All contribute to reporting
 - Beyond required: results, papers, ppt, ...

Text Analysis


- Determine & Design in Y1 → start in Y2
- Criteria for selection:
 - Synergistic with linguistic core & MT
 - Of interest to Government
- Candidates:
 - Geo-referencing (locations, etc.)
 - Summarization
 - Sentiment analysis

Year 1: 10-2010 to 09-2011


Task 1

-  Determine feature sets (morpho, syntactic, functional semantic), and analyze first two languages, writing first set of Linguistic Core translation rules

Task 2


-  Develop and test the tree-transduction method and design additional operators (adjunction, unification) on rich feature structures. Develop rule interpreter. Design Bayesian methods for rule enrichment and constraint learning

Task 3


-  Develop and test end-to-end Linguistic-Core MT and set up evaluation framework based on phenomena. Design relational extraction component
-

Year 2: 10-2011 to 09-2012


Task 1

-  Finalize feature sets, complete linguistic core for first two languages, start work on additional languages

Task 2


-  Complete tree-transduction algorithms and additional operators. Focus on proactive elicitation to enrich and expand coverage beyond linguistic core

Task 3


-  Complete end-to-end MT engine prototype and develop TA engine first for relational extraction and then for summarization

Year 3: 10-2012 to 07-2013


Task 1

-  Develop linguistic core of additional languages, exploit language-family connections as appropriate, expand and refine feature set including semantic annotation

Task 2


-  Refine and improve algorithms with additional operators, such as approximate feature mapping and extended transduction

Task 3


-  Refine and generalize end-to-end MT and TA prototypes, using the linguistic core of the new languages to test for generality. Evaluate engines

Year 4: 10-2013 to 09-2014

Task 1

-  Develop refinements, improvements and enrichments to linguistic core based on re-analysis, complete linguistic core for additional languages

Task 2


-  Develop refinements, improvements and improved robustness of algorithms across the board, including new operators as needed

Task 3


-  Incorporate improvements into MT and TA engines, re-run end-to-end evaluation

Year 5: 10-2014 to 07-2015


Task 1

-  Finalize the enriched feature representations and linguistic cores for all selected languages, write a primer for developing new linguistic cores. Improve/refine linguistic cores for each language

Task 2

-  Finalize transduction and other core algorithms including additional operators. Finalize the enrichment algorithms for proactive elicitation and Bayesian constraint learning

Task 3

-  Finalize and evaluate end-to-end MT on each language, report detailed evaluation of which linguistic phenomena and divergences are handled correctly and how often