

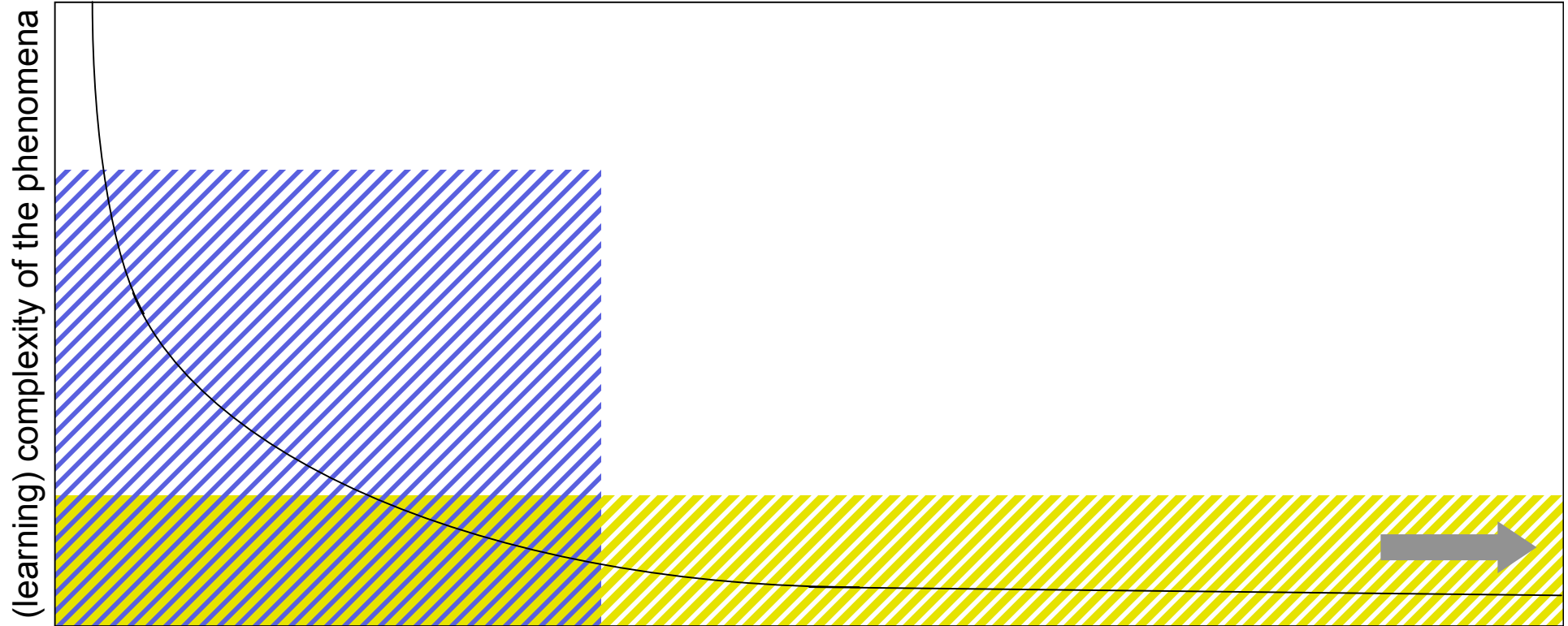
The Core Issues



- ☆ Can everything needed for MT be learned from data?
 - problem 1 -- learning complex and nonlocal regularities
 - problem 2 -- shortage of data both in quantity and quality

- ☆ Can those things that are not learnable, be obtained from RBMT?
 - problem 1 -- Are the limitations of RBMT compensated by SMT?
 - problem 2 -- Can RBMT systems built for all language pairs?

Hypothesis



Linguistic phenomena ordered by their (learning) complexity



linguistic NLP



statistical NLP



Answers



- ☆ Can everything needed for MT be learned from data? **no**
 - problem 1 -- learning complex and nonlocal regularities **not anytime soon**
 - problem 2 -- shortage of data both in quantity and quality **not solved soon**

- ☆ Can the things that are not learnable, be obtained from RBMT? **not all but many**
 - problem 1 Are the limitations of RBMT compensated by SMT? **not all but many**
 - problem 2 Can RBMT systems built for all language pairs? **the price is o.k.**

Predictions



- ☆ My prediction for the distant future I will not share
- ☆ For the near and medium future, I predict a tight race between hybrid MT and SMT including more knowledge sources
- ☆ I believe that hybrid approaches to MT will be leading in applications requiring higher quality and in specialized systems
- ☆ Whereas enriched SMT systems will be leading in inbound informative translation and in language pairs without RBMT
- ☆ Other statistical paradigms combining multiple knowledge sources such as Context-Based MT will be successful for language pairs lacking sufficient volumes of parallel data
- ☆ Finally, I predict that in the future we will have much research on something we may call situated MT, systems that know or guess the exact context and behave quite different for different contexts.