

Machine Translation

Symbolic Methods

Martin Kay

*Stanford University and
The University of the Saarland*

Abstraction

Elimination of

- Special cases
- Exceptions

Spelling rules

Punctuation

Declensions

Conjugations

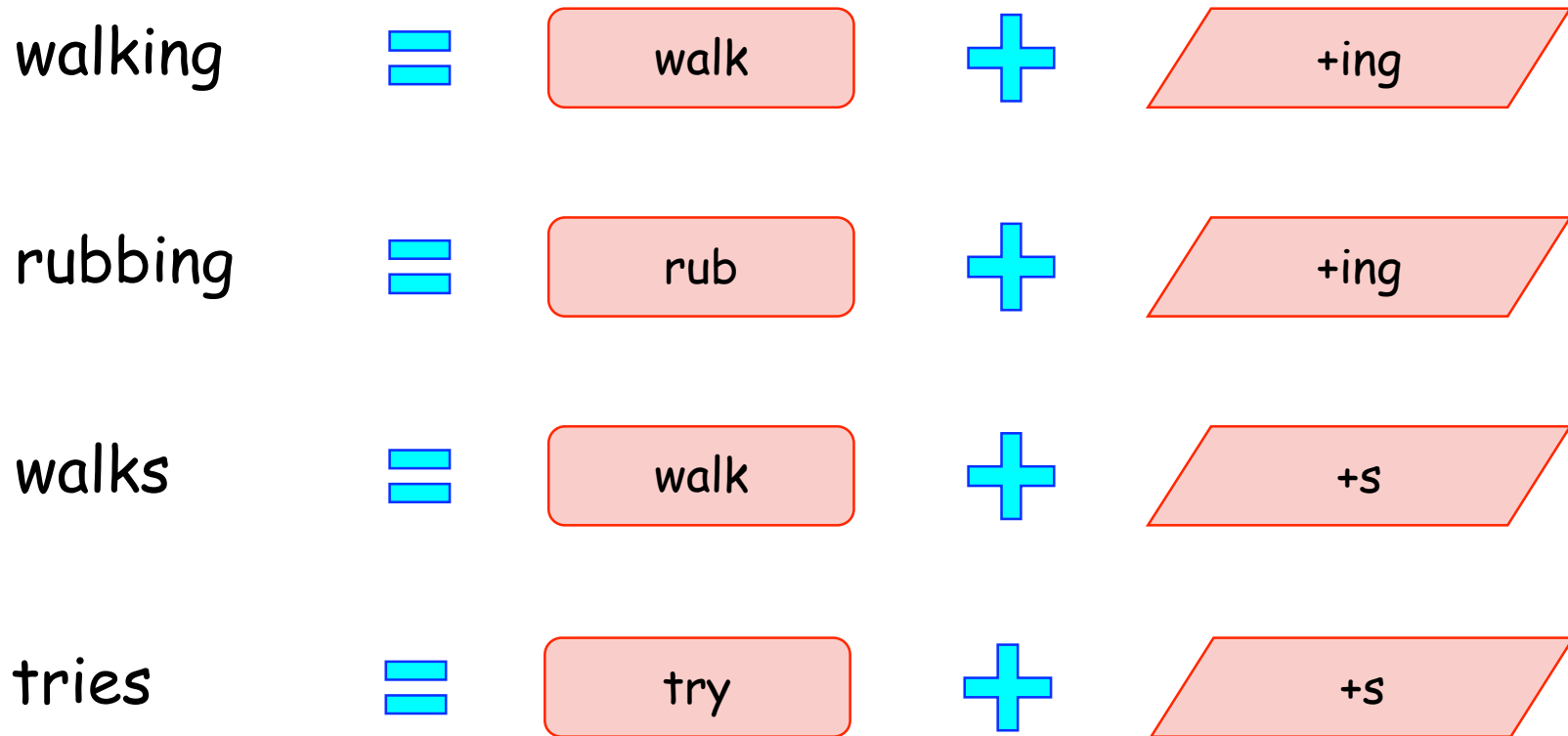
Cases

Prepositions

Moods

...

Morphographemic Abstraction



Spelling idiosyncracies no longer matter

no longer get in the way

Morphographemics

Kind

love

run

manger

try

tie

medico

arco

Kinder

loves

runs

mange

trying

tying

Kindern

loving

running

mangeons

tries

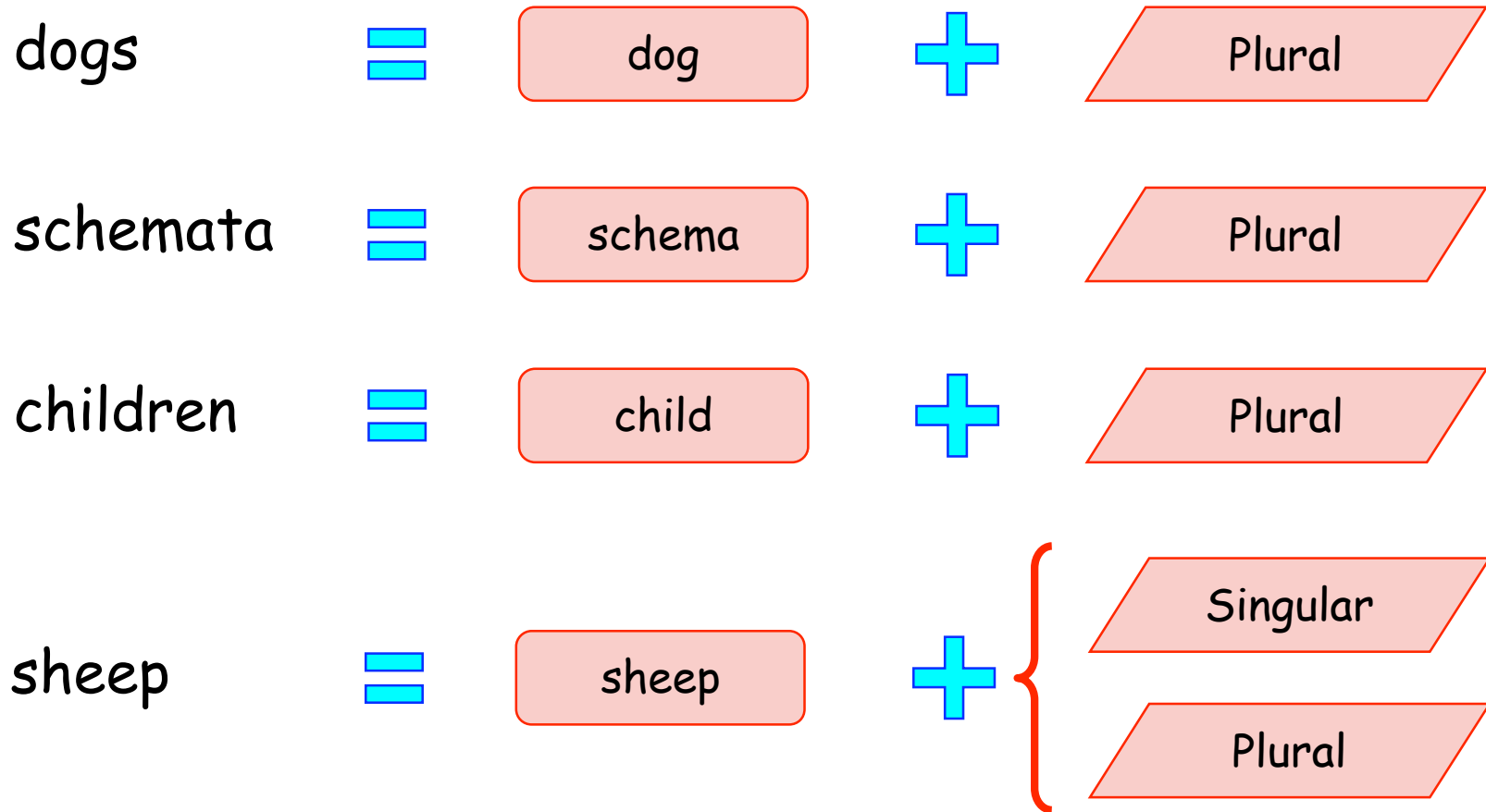
ties

medici

arche

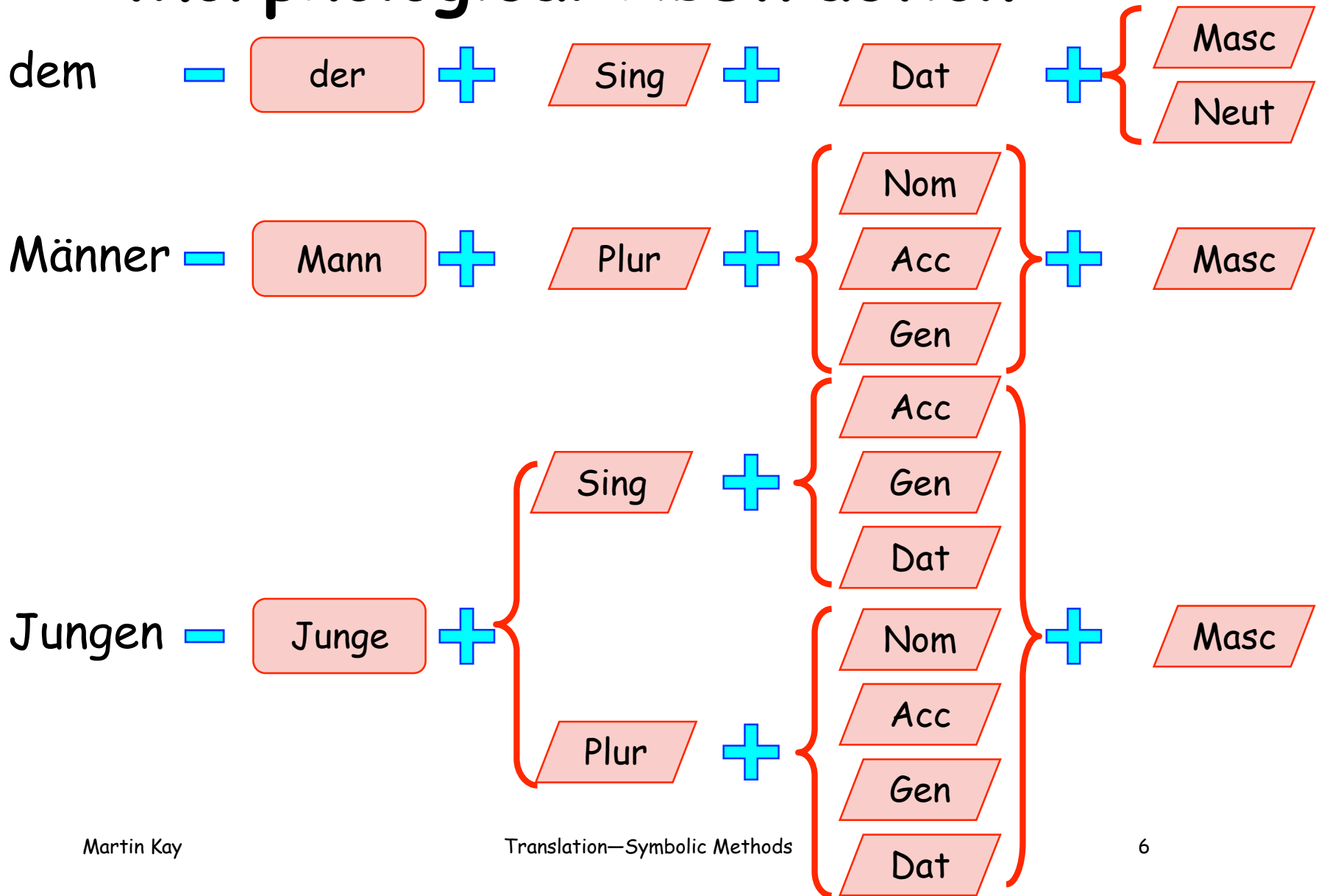
Diacritics

Morphological Abstraction

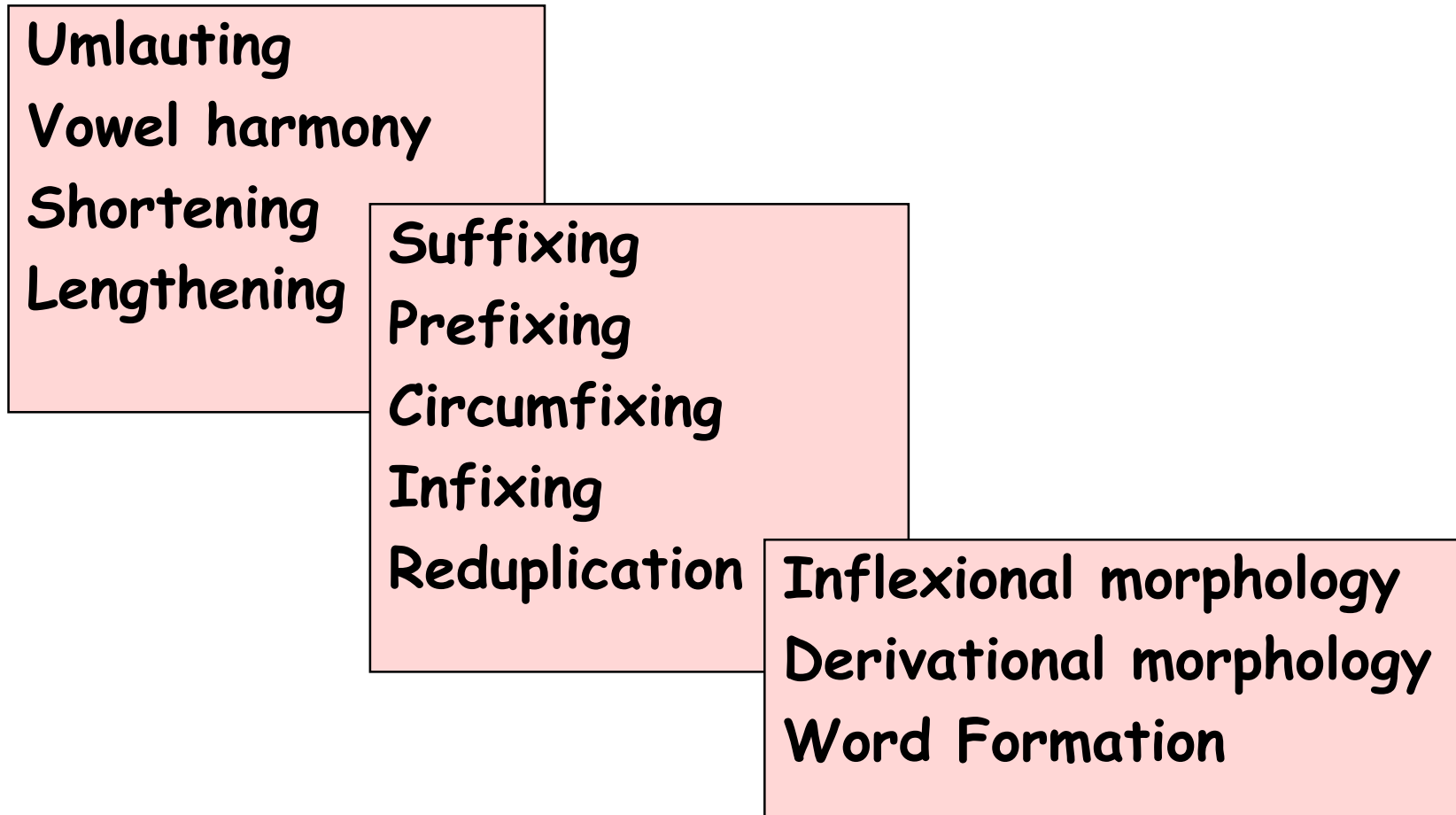


Paradigms and exceptions no longer matter

Morphological Abstraction



Word-level Processes



Syntactic Abstraction

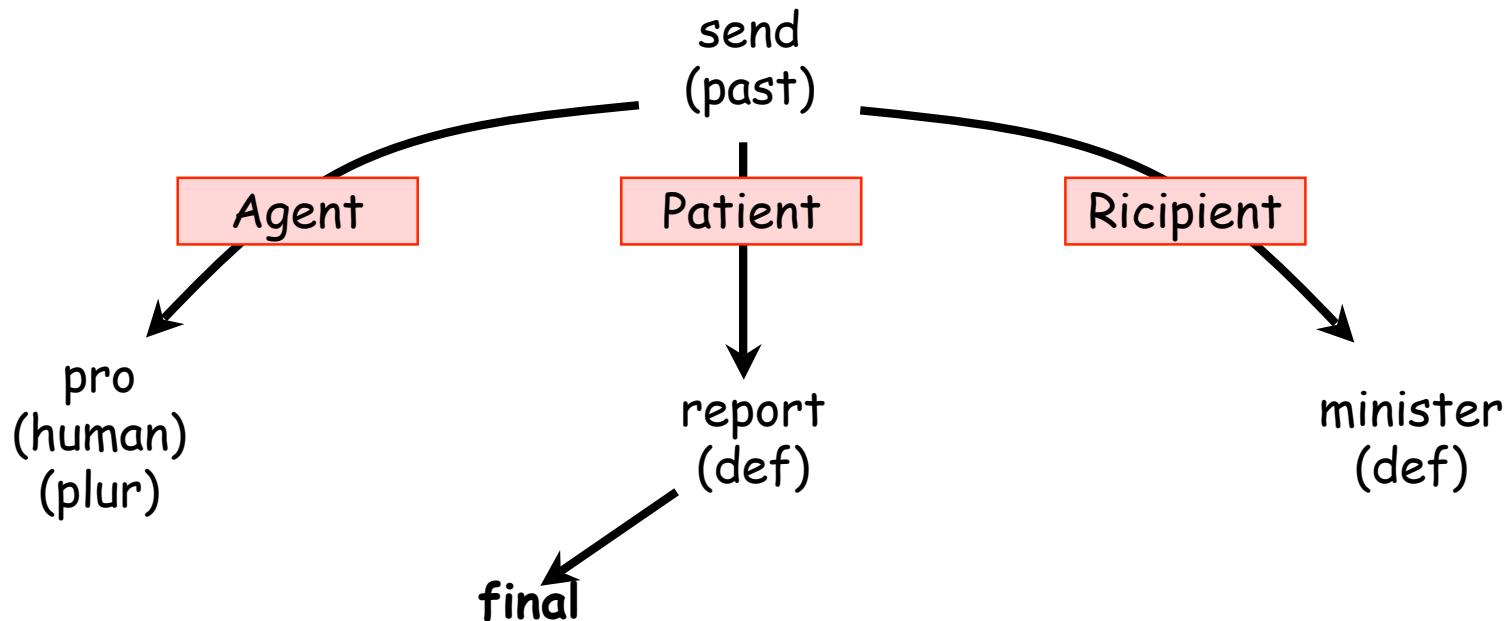
They sent the final report to the minister

They sent the minister the final report

The final report, they sent to the minister

To the minister they sent the final report

The final report was sent to the minister (by them)



Syntactic Abstraction

How much abstraction is enough/too much?

Information structure

John gave this perfect stranger a lot of money

John gave a lot of money to this perfect stranger

Broccoli, I cannot stand!

One thing I cannot stand is broccoli.

It is Ivan that caused all the trouble in the first place.

The more broccoli there is, the less I like it.

Topicalization

What does it *mean* in English/German?

Other Levels

His clever brother always stood in his light

Er stand immer im schatten seines klugen
Bruders

He will not be here until Monday

Er wird erst Montag da sein

Cela vous plait?

Do you like that?

Hans schwimmt gern

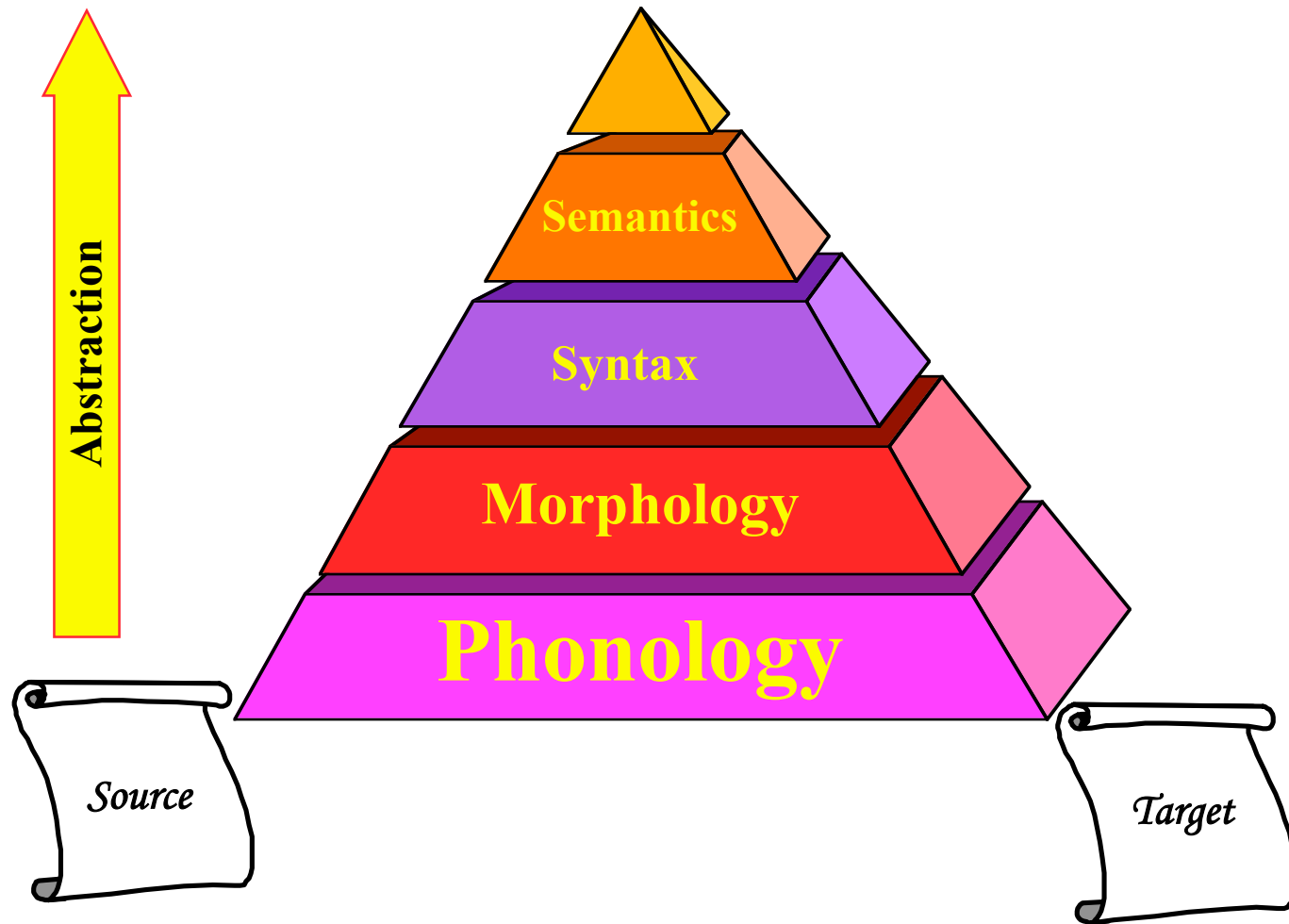
Hans likes swimming/to swim

Syntax? — Adjective order

Opinion	Size	Age	Shape	Color	Origin	Material	Purpose	
Fine	big	old				wooden	storage	boxes
	little			blue	Mexican			model
Funny			round				meeting	room
					farm	vegetable		product

How to classify
organic
recursive
soft
running
... ?

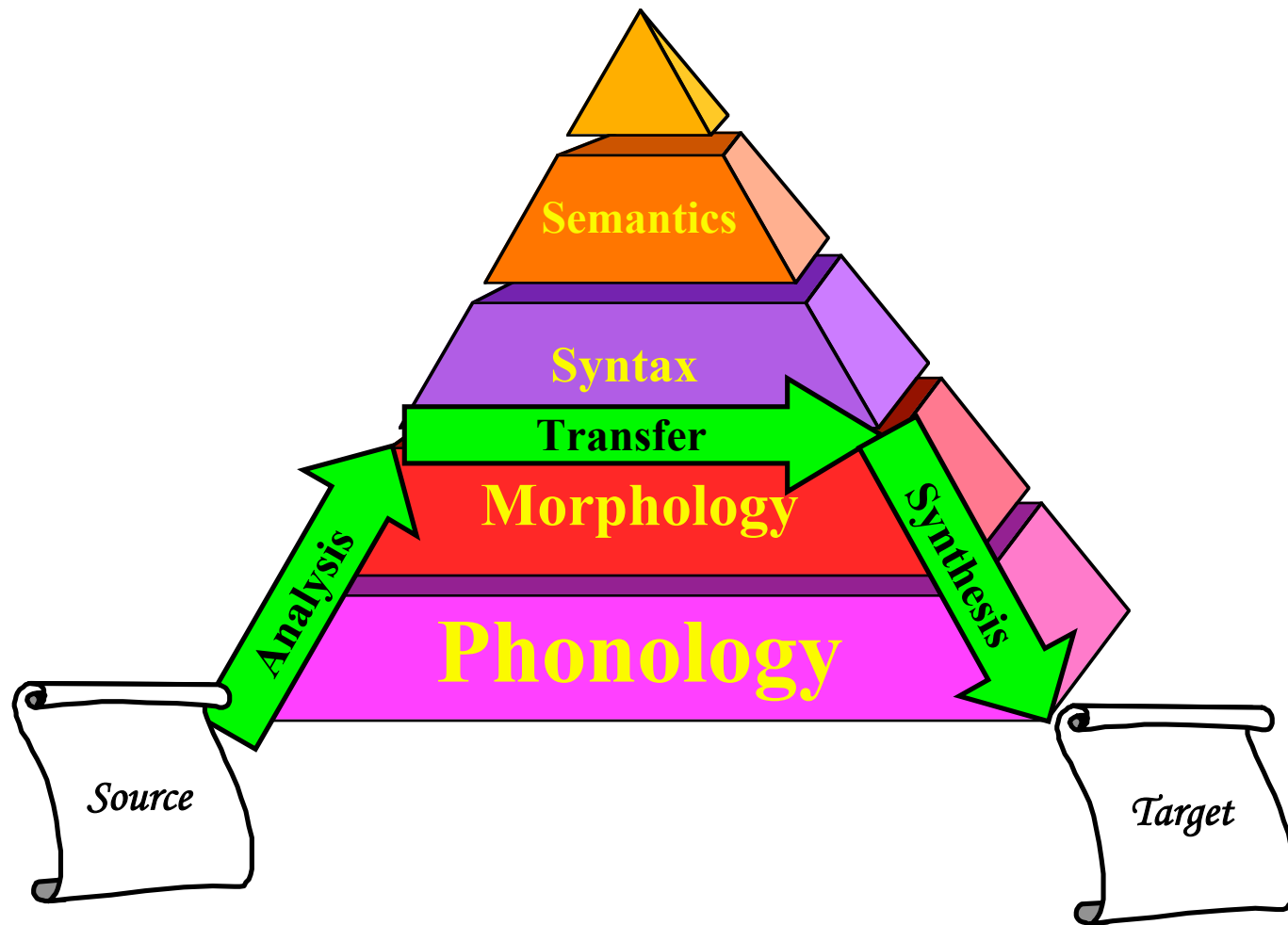
The Vauquois Triangle



The Transfer Approach

Analyze to some level of abstraction L
Transfer
Generate

The Vauquois Triangle



Commercial Systems

Do not follow the model closely:

—Levels of abstraction are

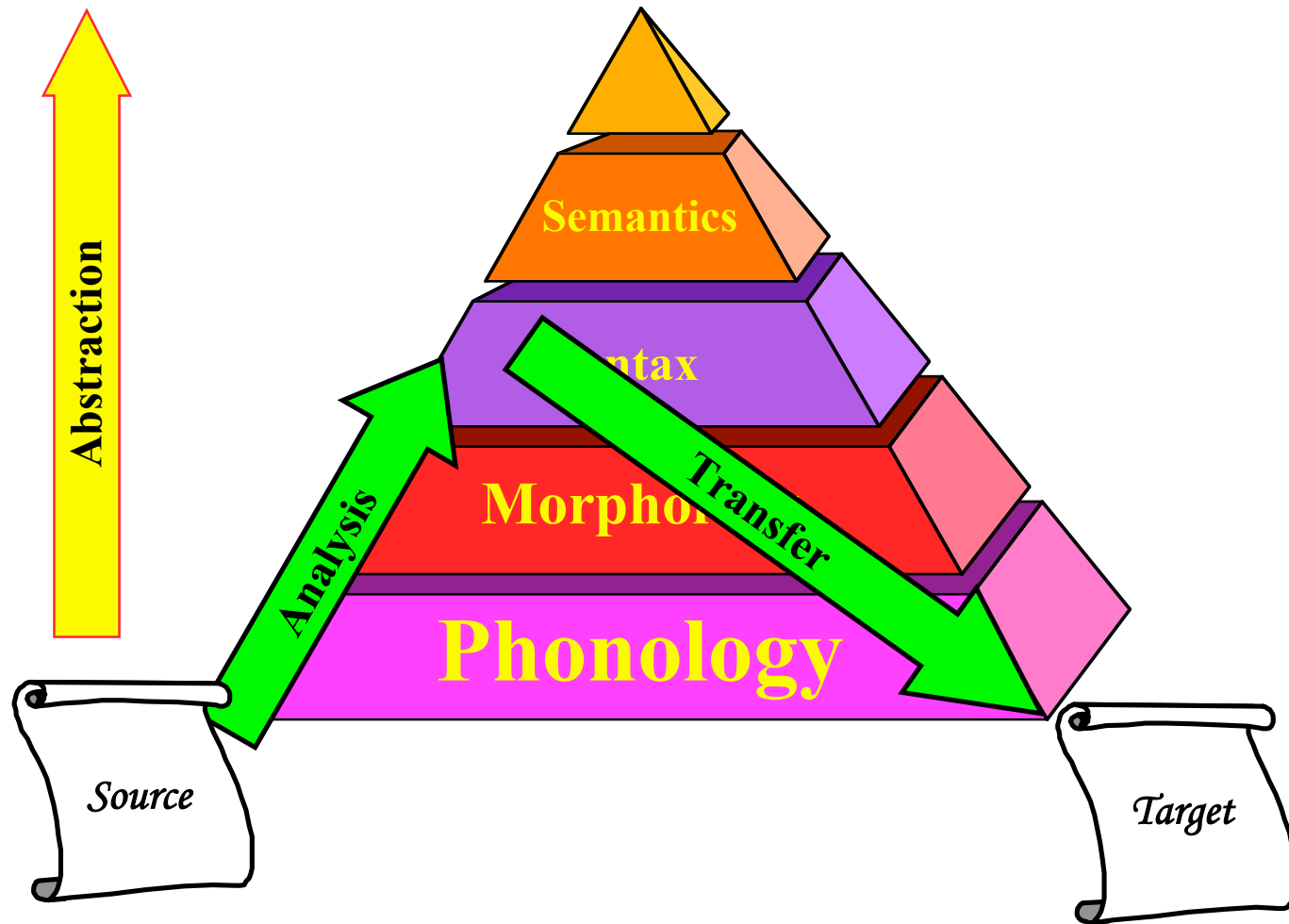
- Not strongly separated
- Are weakly formalized at best

—Generation Levels are largely eliminated

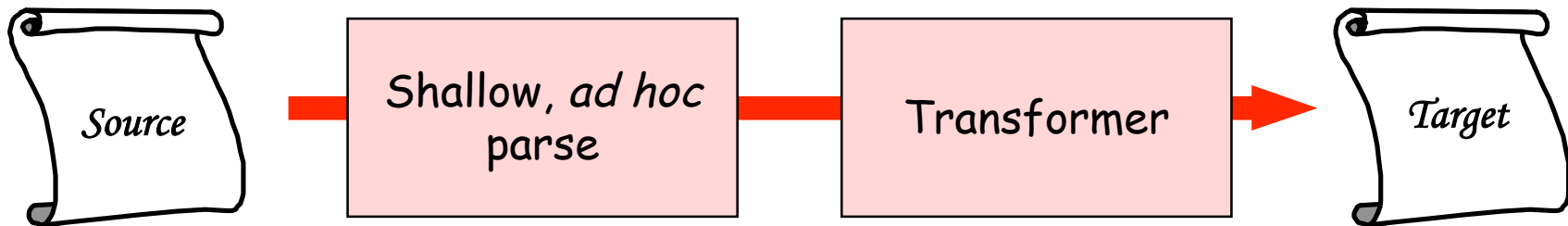
Are almost entirely deterministic

Aim for speed

The Vauquois Triangle



The Standard Approach



Commercial Systems

Rely on

- Tuning the lexicon to the domain
- Huge inventories of set phrases
- Selectional restrictions

Commercial Systems

- **Weak points**
 - Early binding — no nondeterminism
 - Result will, in general, be ungrammatical

The Standard Approach

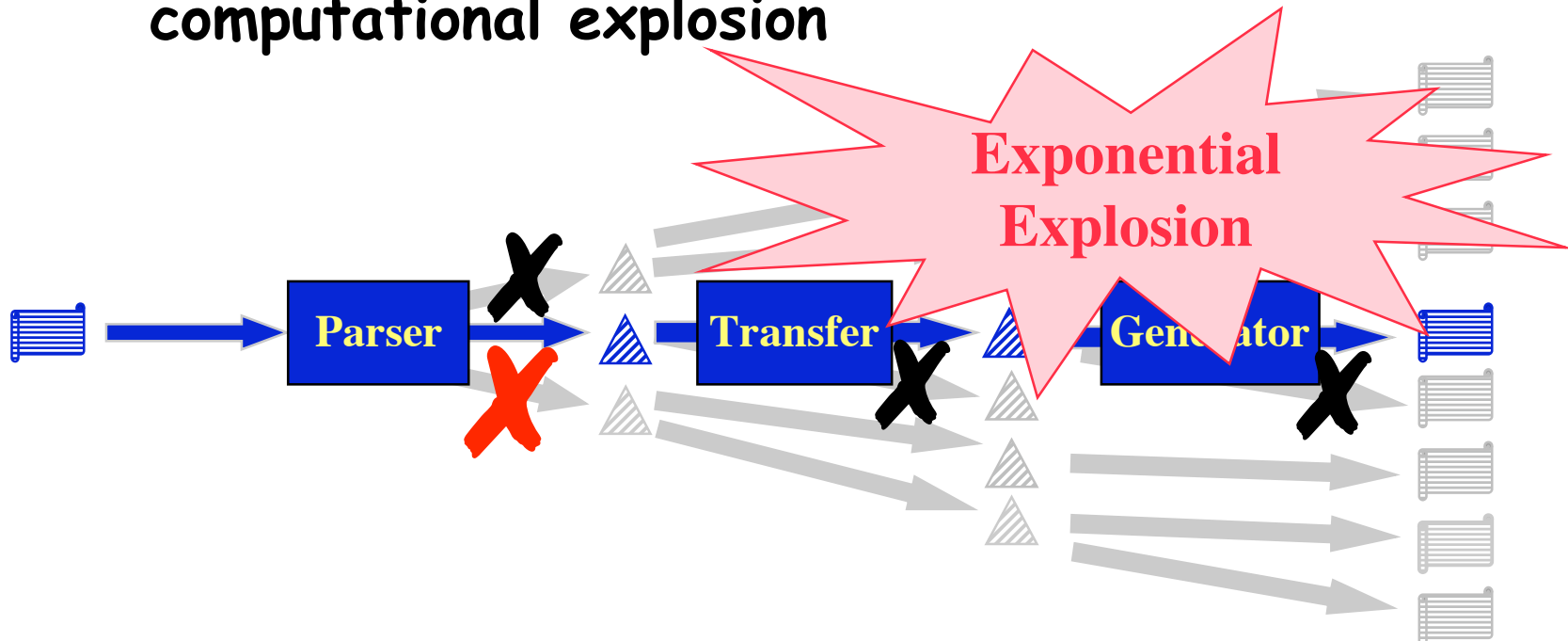
Separate modules for simplicity,
maintainability, reuse



The Standard Approach

Separate modules for simplicity,
maintainability, reuse

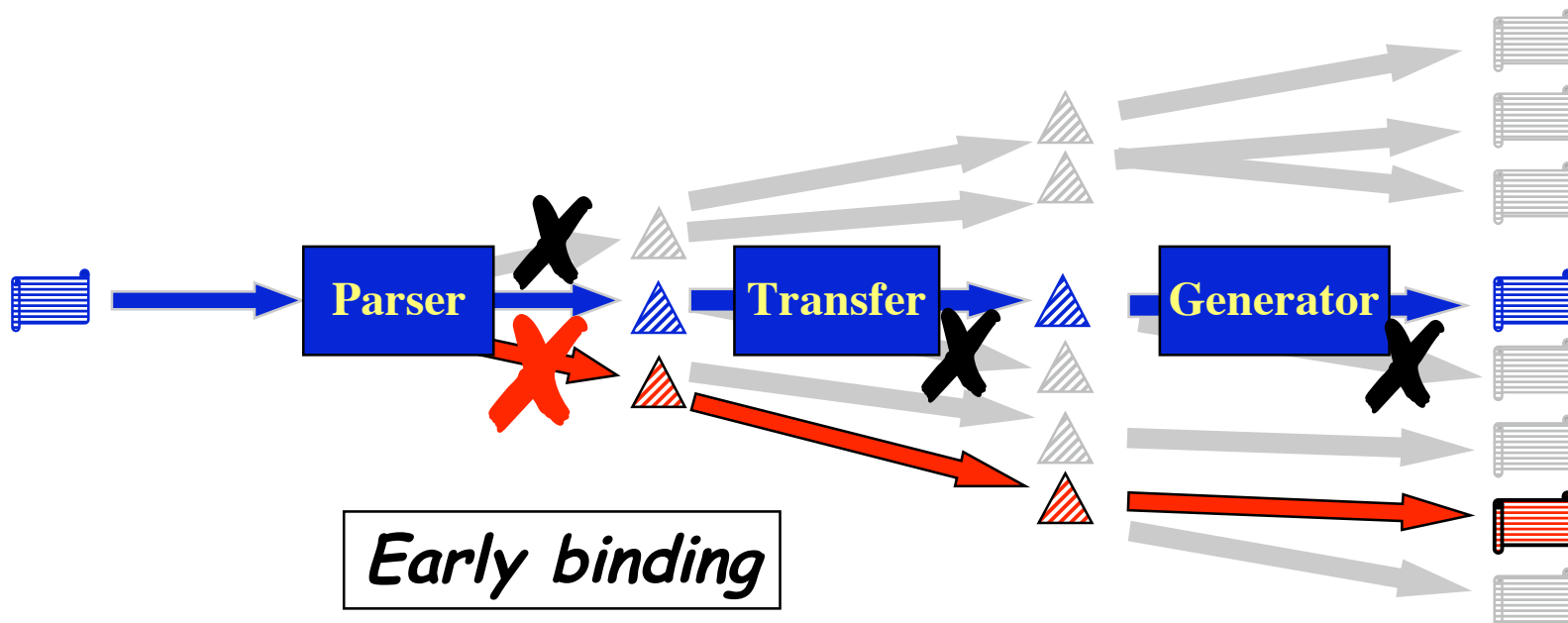
Heuristic filters are applied early to avoid
computational explosion



The Standard Approach

Separate modules for simplicity,
maintainability, reuse

Heuristic filters are applied early to avoid
computational explosion



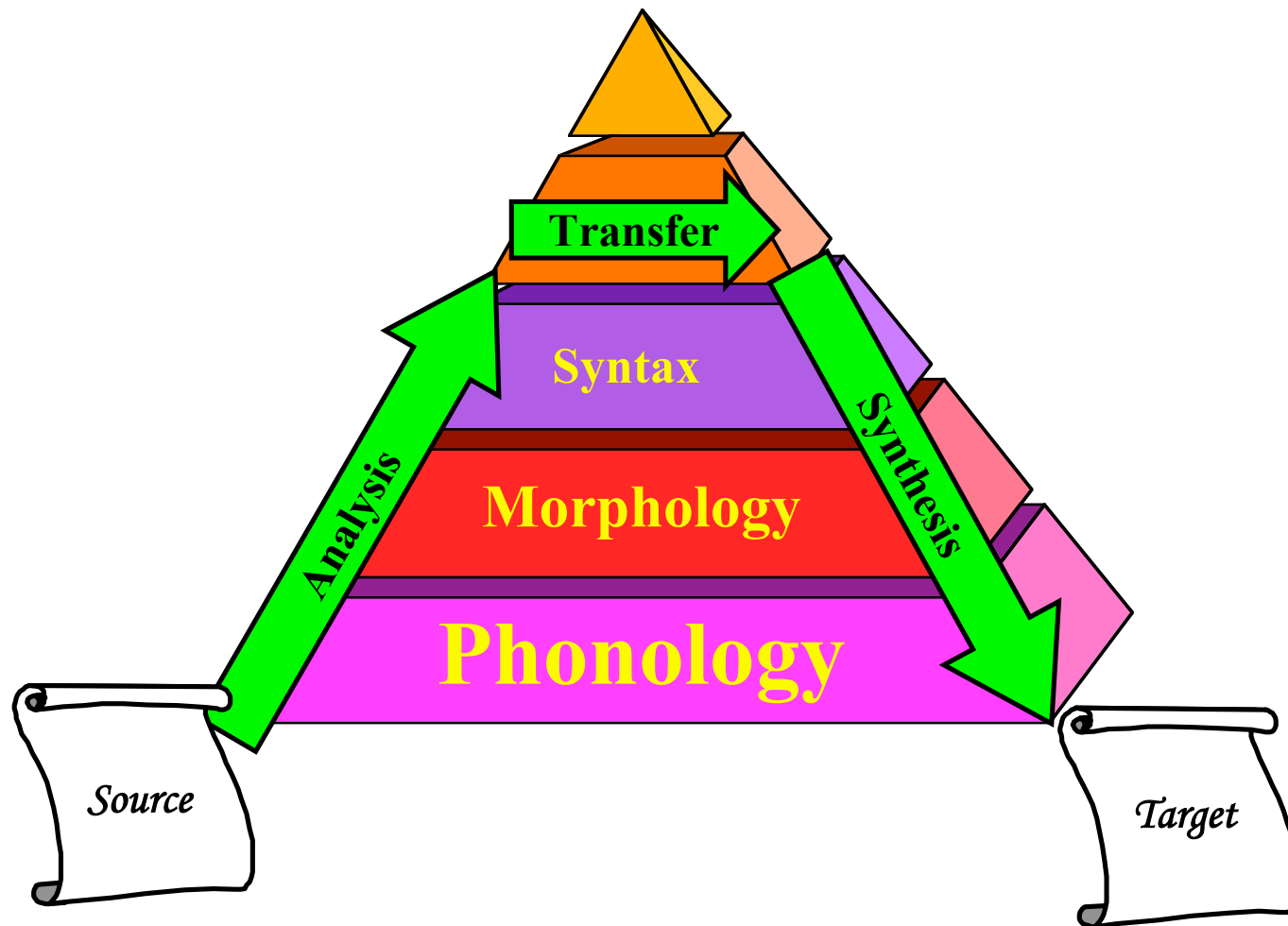
Assessment of the Standard Approach

- Robust
- Can produce *word salad*
- *Ad hoc* and hard to maintain
- Bilingual and unidirectional

Academic Approaches

- More abstraction — appeal to AI
- Equal weight to analysis and generation
- Formalisation
- Avoid early binding

Academic Approaches



Academic Approaches

Problems

Time

Robstness

Ambiguity

Linguistics

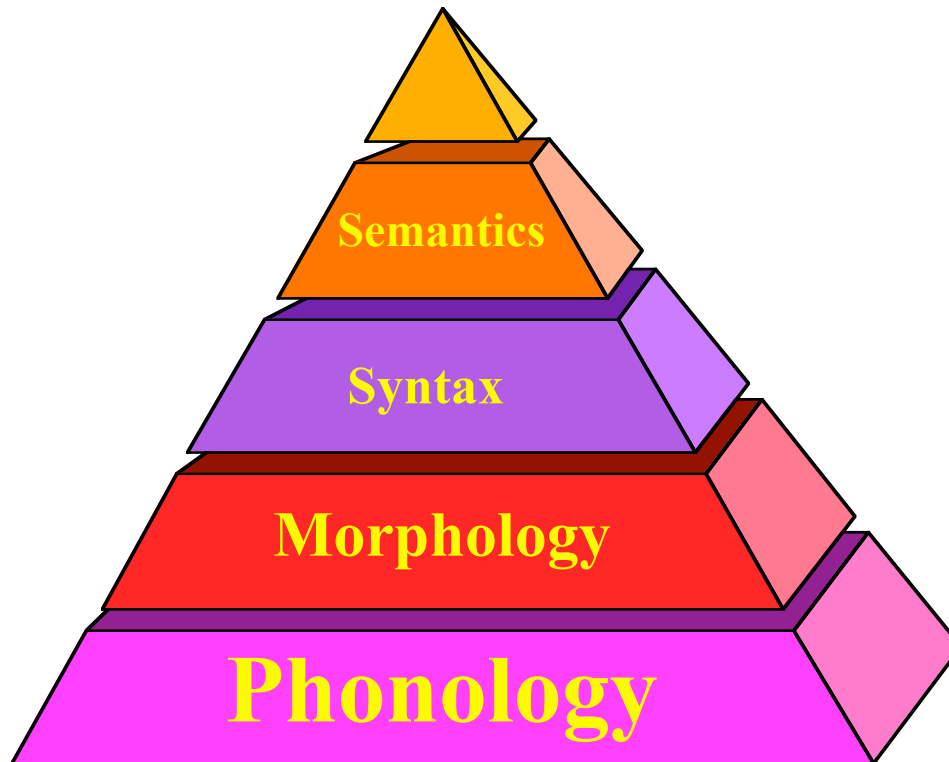
Can identify

Ambiguity

But not resolve

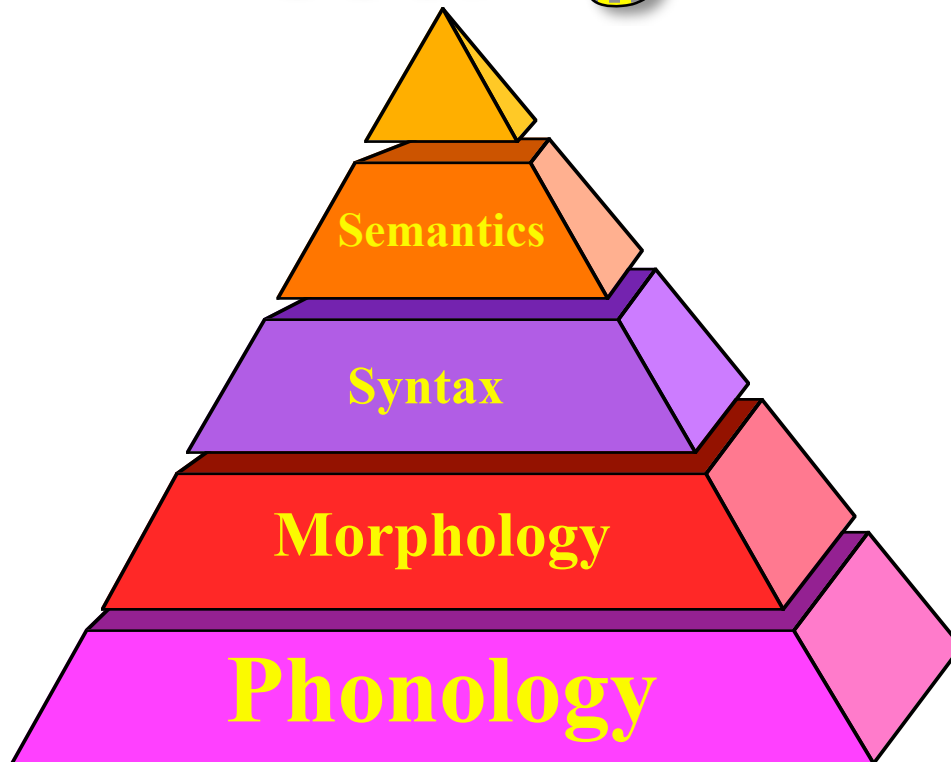
The Vauquois Triangle

What is this?



The Vauquois Triangle

Interlingua



If you abstract enough

You will be left with Pure Thought

OK. So what is wrong with that?

Interlingua must

- Represent whatever any language can represent, even if it will often be lost in translation.
- Problems of (non)overlap in the semantic grid.