

MT Summit XI

Workshop on Using Corpora in Natural Language
Generation and Machine Translation

September 11, 2007

Is NLG still Relevant to MT?

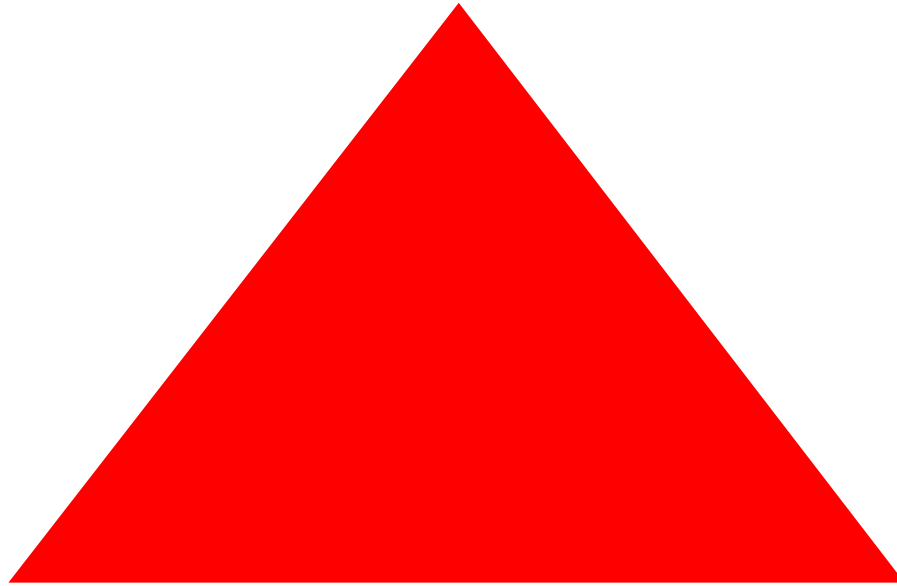
Nizar Habash

Columbia University

Center for Computational Learning Systems

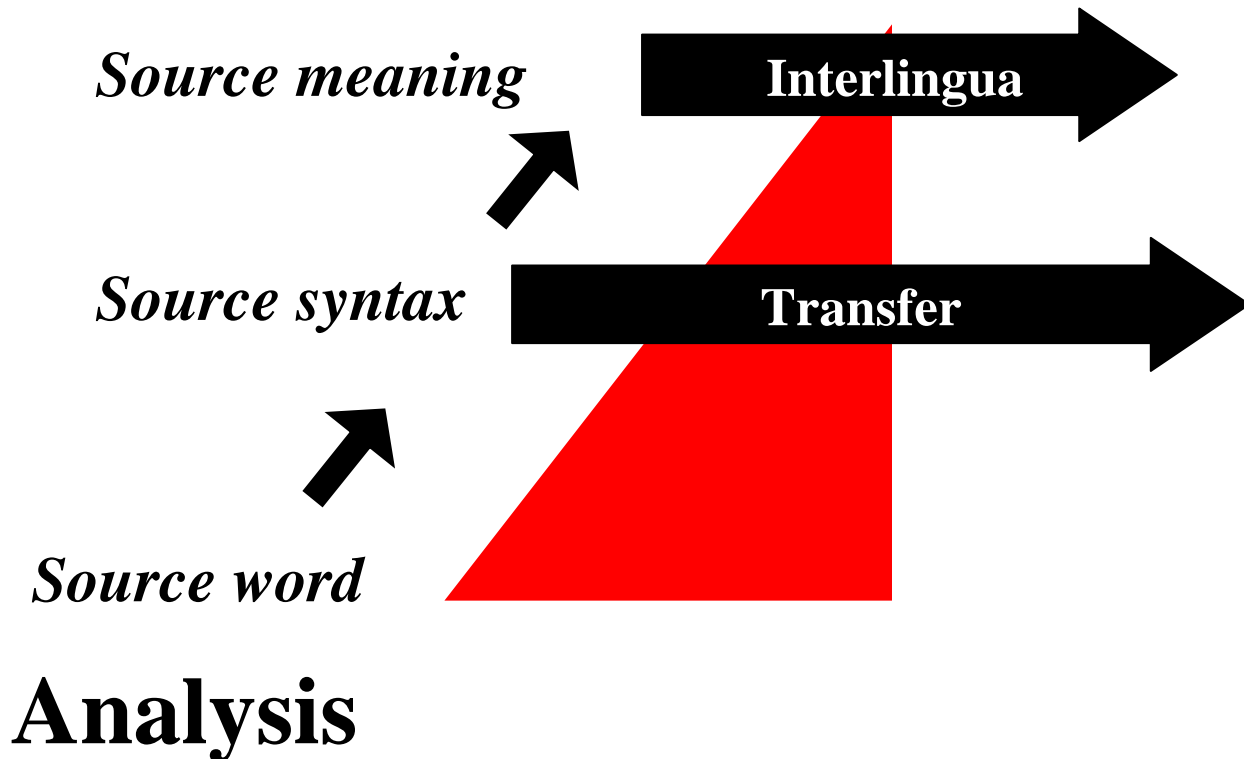
Once upon a time...

- In the olden days of Rule-based MT
- Natural language generation had a very prominent role...



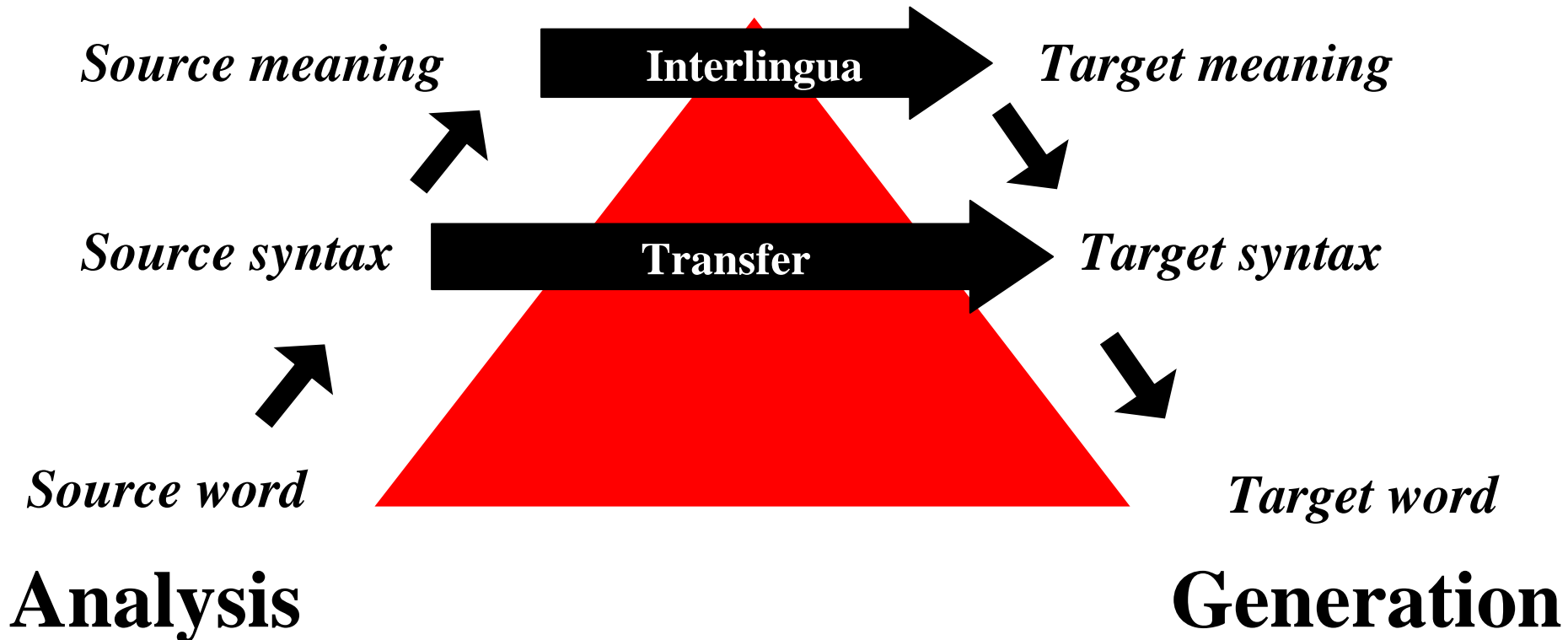
Once upon a time...

- NLG was HALF of the MT story...



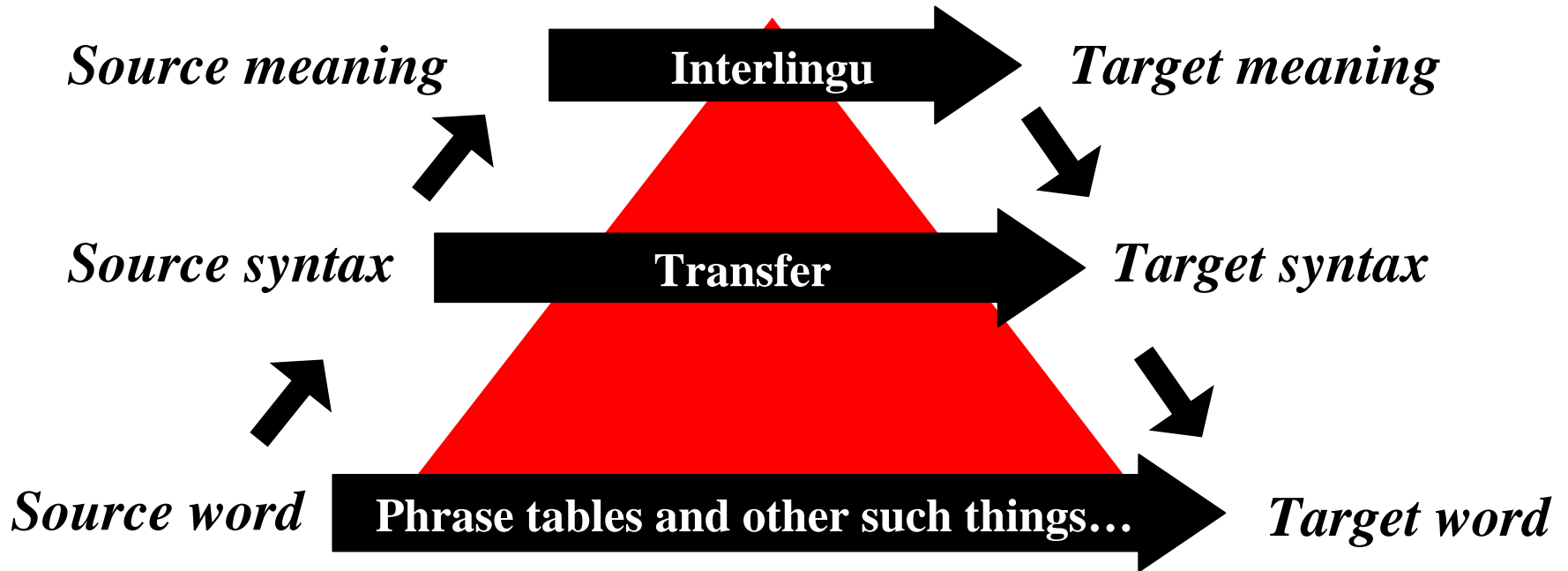
Once upon a time...

- In those days, we had something to generate from...



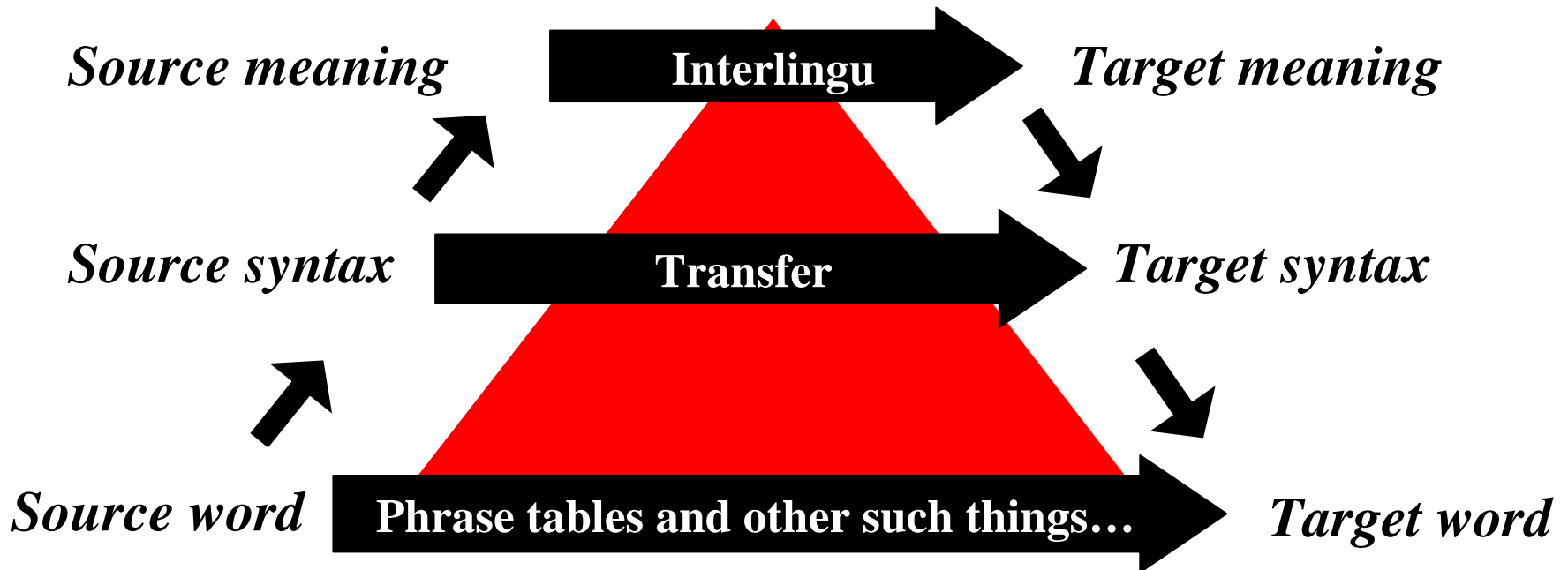
Once upon a time...

- Then came Statistical MT...



Once upon a time...

- Then came Statistical MT...



Once upon a time...

- Then came Statistical MT...
... and NLG as we knew it became irrelevant

Source word

Phrase tables and other such things...

Target word

Why has NLG become irrelevant to MT?

- Deeper levels are hard to model (whine)
- Deeper levels are useless (whine)
- Will NLG help my Bleu? (cheese)
- SMT models the complete process of translation end-to-end
- Generating English is easy. So why bother? (خليني اقوالك ليش)
- Have you seen my MT pyramid, I can't think without it? (whine)

NLG can still help MT! (really)

- Translation from a morphologically poor language into a morphologically rich language will require sophisticated generation
 - English to Arabic, Chinese to Czech, etc.
 - Fill in monolingual gaps
 - case, number & gender agreement, etc.



NLG can still help MT! (really)

- Translation from a morphologically poor language into a morphologically rich language will require sophisticated generation
 - English to Arabic, Chinese to Czech, etc.
 - Fill in monolingual gaps
 - case, number & gender agreement, etc.

“ Three universities were built ”

bnyt <i>'were built'</i>	vIAv <i>'three'</i>	jAmEAt <i>'universities'</i>	jdypd <i>'new'</i>
Fem+Sg	Masc+Sg-Nom	Fem+PL-Gen	Fem+Sg-Gen
Verbs in VSO order are always Sg and agree in gender only	Numbers agrees by gender inversion		Adjectives of plural irrational nouns are fem+Sg

NLG can still help MT! (really)

- Perhaps some NLG components can be used not as intended...
 - For paraphrase expansion of source side
 - Lexical expansion
 - Word-reordering
 - For rescoring translation hypotheses
 - Catch hallucinations and missing content words
 - Document planning for SMT
 - Use NLG to guide some SMT choices

How can I help?

- For NLG to be relevant again in MT, NLG researchers must
 - Learn more about Statistical MT systems to be able to integrate NLG components in them
 - Think outside the pyramid
 - Encourage research on translation into language that require sophisticated NLG solutions
 - Morphologically rich target languages
 - Different genres and domains where linguistic generalizations are useful

Thank you!

Nizar Habash

habash@cs.columbia.edu

Columbia University