

CLARIFY: Human-Powered Training of SMT Models

D. Scott Appling & Ellen Yi-Luen Do
darren.scott.appling@gatech.edu; ellendo@cc.gatech.edu



Overview

- Statistical Machine Translation?
- CLARIFY Framework
- Human-Powered Computing
- Developing Visual Data
- Games for Translation

Statistical Machine Translation (SMT)

- Produce the best semantic translation
- Advantages
 - Cheaper < Human
 - Cross-Cultural Communication
 - Mobility



SMT Approach

The diagram shows the SMT equation with several annotations and arrows:

- target sentence**: An arrow points from this text to the \mathbf{f} in the denominator of the probability $p(\mathbf{e}|\mathbf{f})$.
- source sentence**: An arrow points from this text to the \mathbf{e} in the numerator of the probability $p(\mathbf{e}|\mathbf{f})$.
- not calculated in practice**: An arrow points from this text to the $\frac{1}{Z}$ term in the equation.
- model weights**: An arrow points from this text to the λ_i term in the summation.
- feature functions**: An arrow points from this text to the $h_i(\mathbf{e}, \mathbf{f})$ term in the summation. Below this text, it says "e.g. language model score, forward translation score".

$$p(\mathbf{e}|\mathbf{f}) = \frac{1}{Z} \exp \sum_{i=1}^n \lambda_i h_i(\mathbf{e}, \mathbf{f})$$

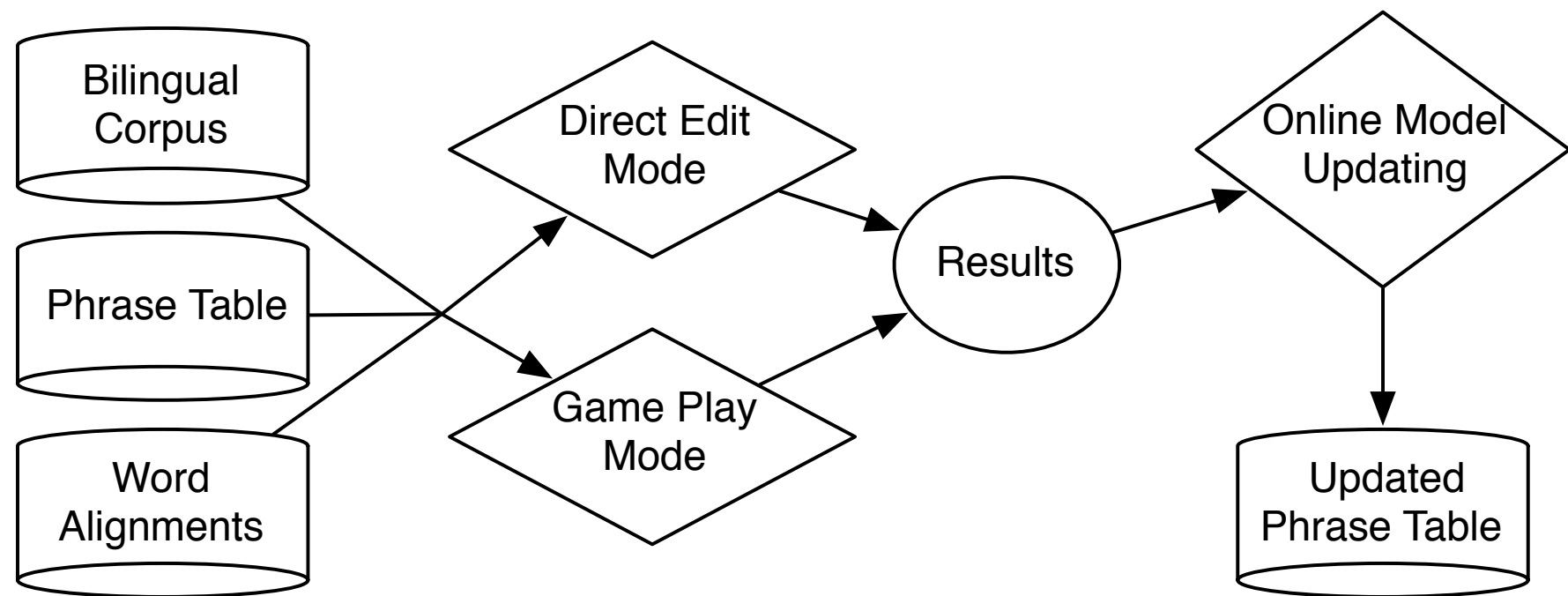
(Koehn et. al., 2007)

What is CLARIFY?

- Data Acquisition Framework
- Corpus Annotation Tools
- Interactive Games
- Paraphrase Validation



CLARIFY System Process



Human-Powered Computing

- Learning from Humans

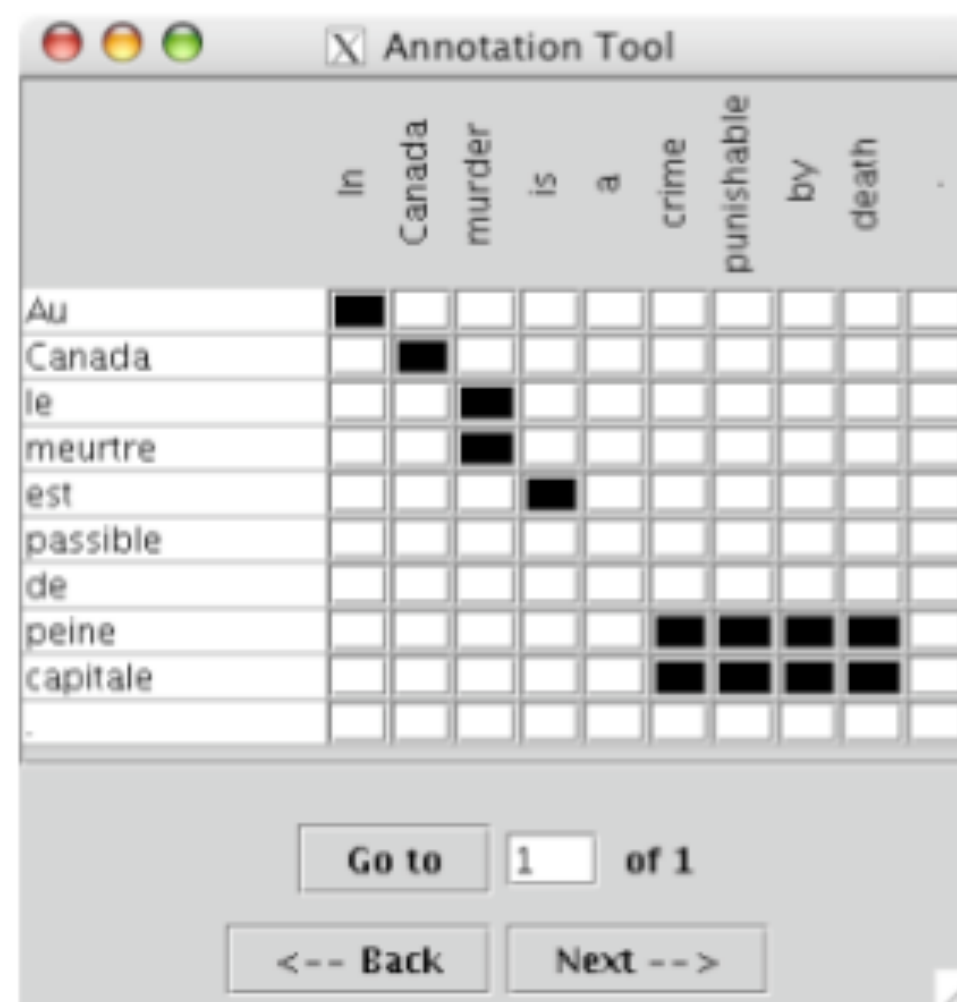
- Training Data

- *The ESP Game*
(von Ahn & Dabbish, 2004)



Visual Data Editors

Improved Statistical Translation Through Editing
(Callison-Burch et. al., 2004)



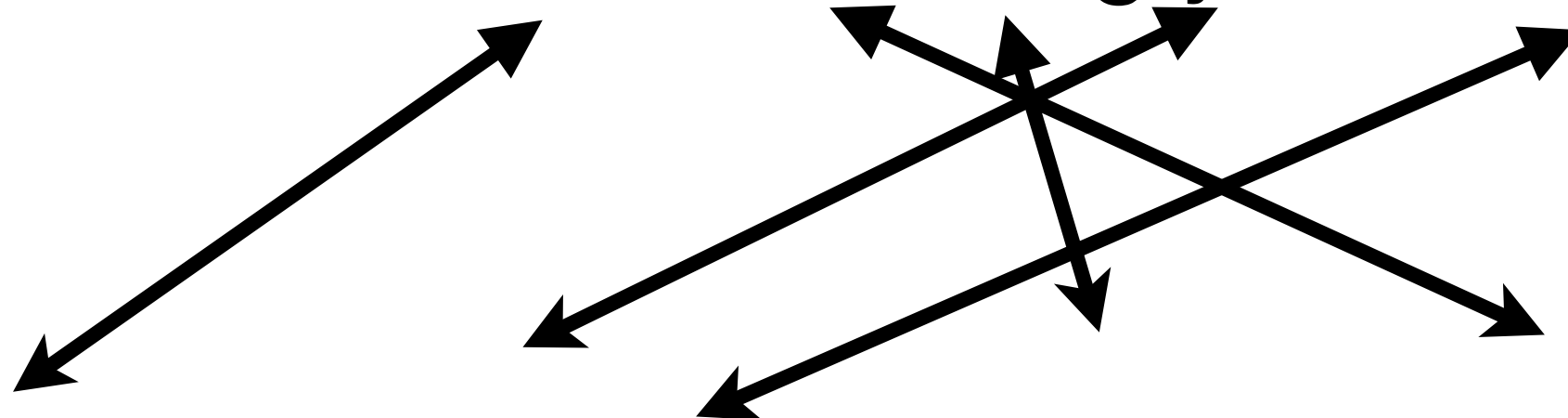
CLARIFY Knowledge Editors

- Visually Editing Machine Translation Model Information
 - Word Alignments
 - Phrase Alignments
 - Paraphrasing

Word Alignment Editor

The cat liked eating yellow bugs .

猫 は 黄色い 虫 が 食べる事 が 好き だった。



Untitled

Clarify Knowledge Modeler

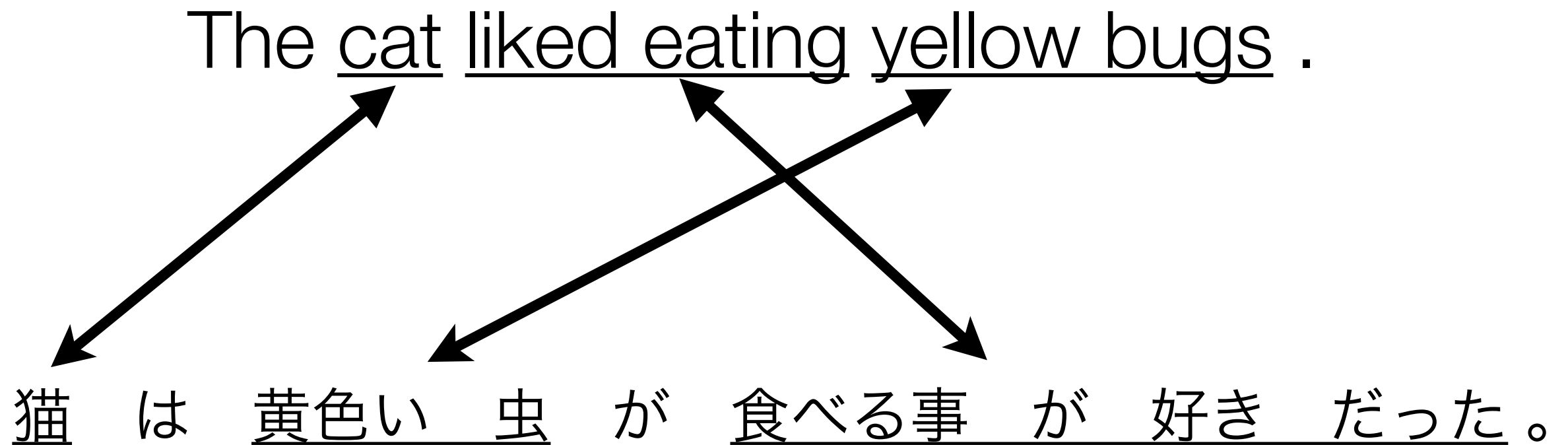
Direct Edit ▼

Load Model Word Alignment Phrase Alignment Paraphrasing Results

この 小包 を 日本 へ 送り たい の です が 。 NULL

I'd like to send this parcel to Japan. NULL

Phrase Alignment Editor



Clarify Knowledge Modeler

Direct Edit ▼

Load Model

Word Alignment

Phrase Alignment

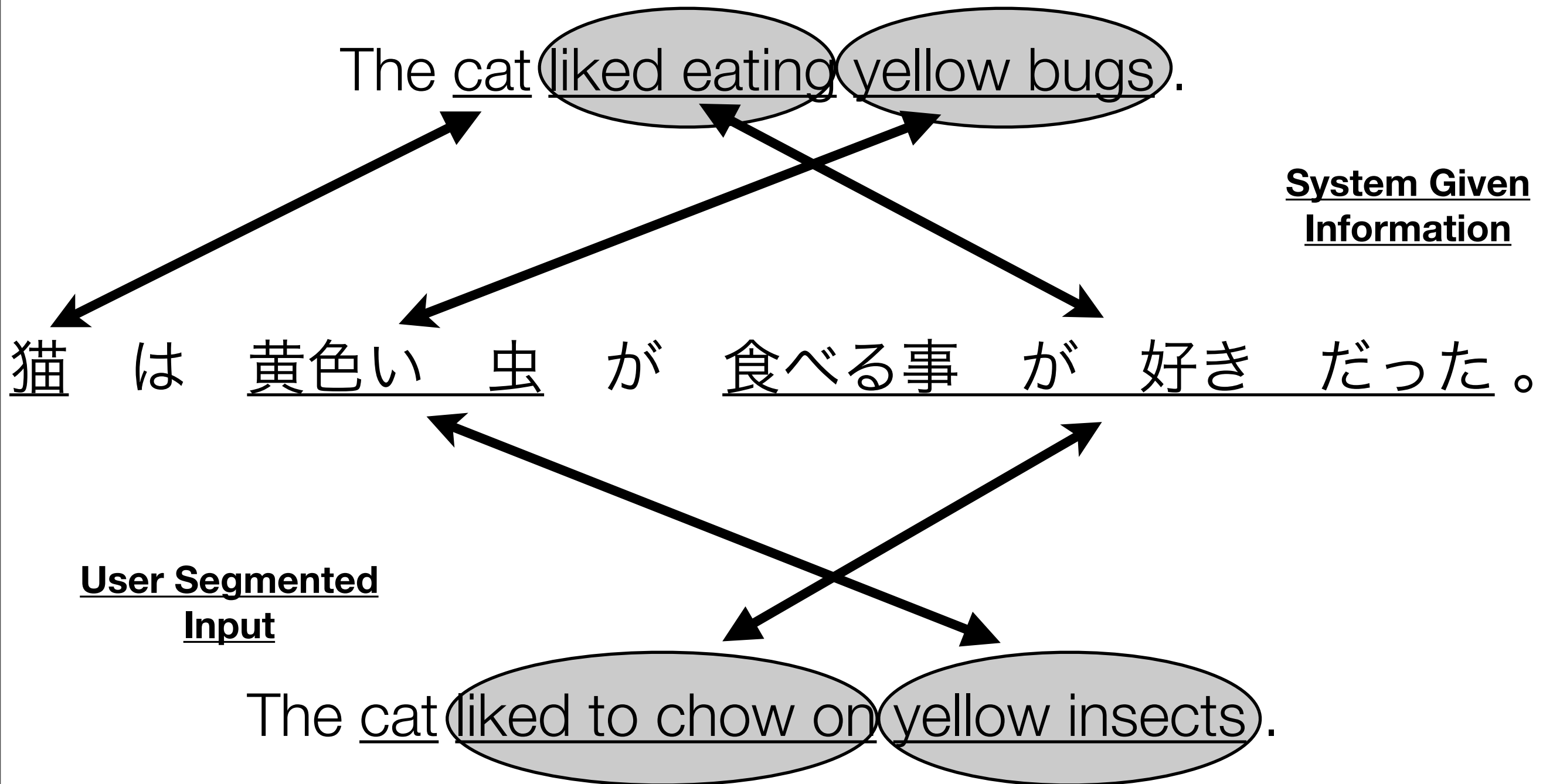
Paraphrasing

Results

この 小包 を 日本 へ 送り たい の です が 。 NULL

I'd like to send this parcel to Japan. NULL

Paraphrasing Editor



Untitled

Clarify Knowledge Modeler

Direct Edit ▼

Load Model Word Alignment Phrase Alignment Paraphrasing Results

Source Sentence

すみません、分かりません。

Paraphrase for: ☒ Source ☐ Target

すみません、知りません。

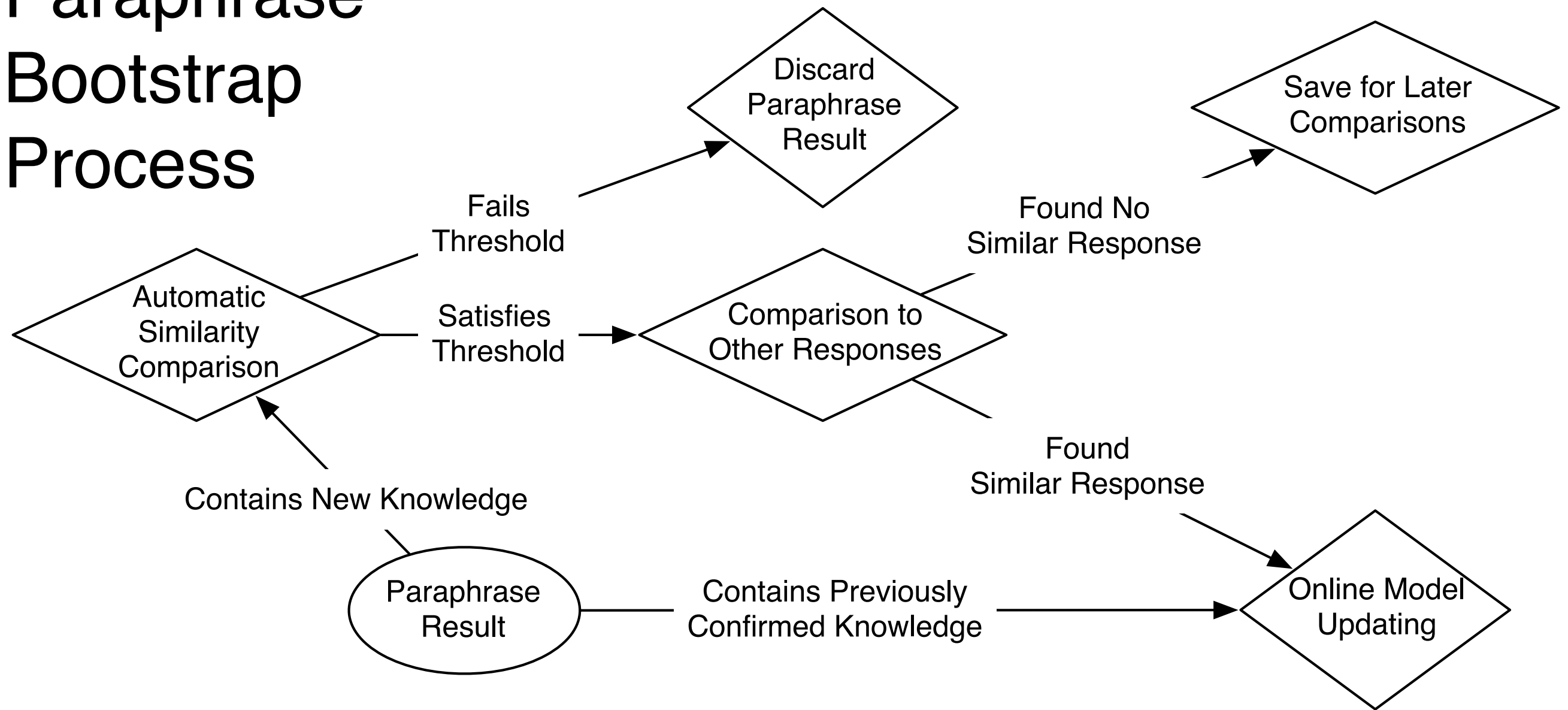
Target Sentence

I'm sorry I don't know.

CLARIFY Translation Games

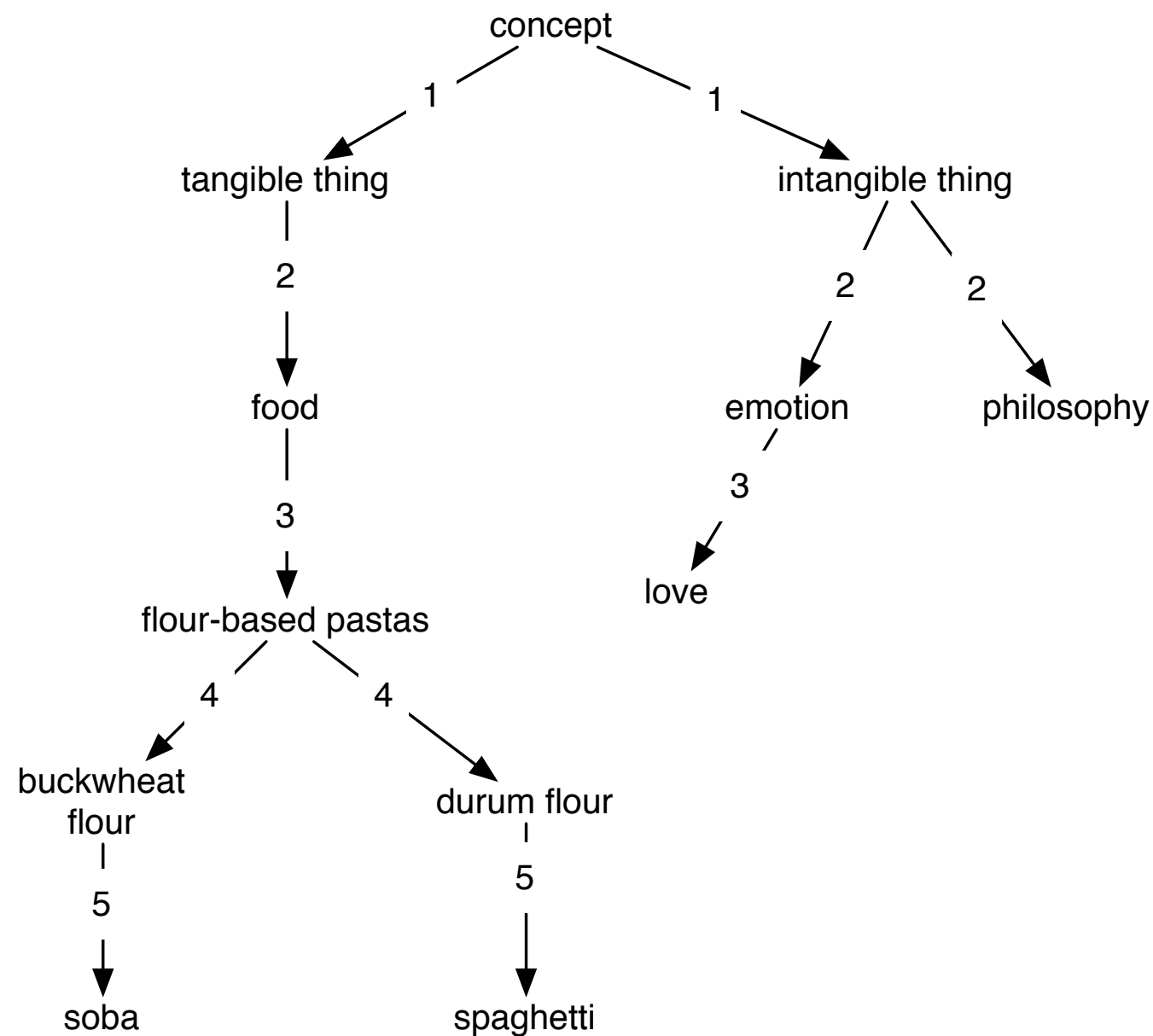
- Employ Three Previous Editors
- Constraints on Games
- Developing Intuitive Game Designs

Paraphrase Bootstrap Process



WordNet-based Similarity

- Deciding on new Paraphrase Acquisition



Accounting for User's Experience

- Users have Different Levels of Knowledge
- Want Smoother Modeling
- Identify User Language Ability
- Knowledge Source Score Component

CLARIFY Review

- Direct Editors
- Interactive Games
- Language Ability Discrepancies
- Paraphrase Knowledge Validation

CLARIFY: Human-Powered Training of SMT Models

Thank You!

D. Scott Appling & Ellen Yi-Luen Do
darren.scott.appling@gatech.edu; ellendo@cc.gatech.edu

