

# C-TOOLS Automated Grading for Online Concept Maps Works Well with a Little Help from “WORDNET”

Scott H. Harrison<sup>1</sup>, Joshua L. Wallace<sup>2</sup>, Diane Ebert-May<sup>3</sup> And Douglas B. Luckie<sup>2,4</sup>

<sup>1</sup>Department of Microbiology and Molecular Genetics, <sup>2</sup>Lyman Briggs School of Science, <sup>3</sup>Department of Plant Biology, and <sup>4</sup>Department of Physiology, Michigan State University

## Abstract

Criterion concept maps developed by instructors in the C-TOOLS project contain numerous expert-generated or “correct” propositions manually created by expert users connecting two object or subject phrases together with a linking phrase. WordNet®, an electronic lexical database, was then used to construct additional proposition derivatives by supplying different linking phrases in place of those originally specified by users. Derived propositions were made from original propositions by substituting linking phrase verbs with troponyms, synonyms, and antonyms. During the past year about 1300 students created concept maps (with 35404 propositions) aided by automatic grading and the WordNet propositions. We studied how successful WordNet was at creating valid additional linking words like those generated by experts. By comparing manual assessments of derived propositions to manual assessments of original propositions, the persistence of correctness was evaluated for determinative factors such as frequencies and senses of usage. Results from data analyses are compared to parameters for WordNet usage in order to better determine the potential for the refining of concept map grading algorithms. An empirical approach to word sense relationships is presented as an iterating step to progressively prototype and test optimizations.

## Examples of Synonyms, Antonyms, Troponyms and Synsets

Original	Nuclear Membrane ? holds ? Nucleus (1)	<i>Manual scorings shown in parentheses</i>
Synonym	Nuclear Membrane ? contains ? Nucleus (1)	
Troponym	Nuclear Membrane ? superannuate ? Nucleus (X)	
Antonym	Plasma Membrane ? let go of ? Nucleus (X)	

WordNet-derived synonyms, antonyms & troponyms were made in sets of 3 independent of internal synset structure. For example:

3 synonyms of “holds” are: {“contains”, “carries”, “prevents”}

⏟
⏟  
 Synset #1                      Synset #2

Synset #1 is valid. Synset #2 is invalid.  
 Nuclear Membrane **contains** Nucleus (1); Nuclear Membrane **carries** Nucleus (1)  
 Nuclear Membrane **prevents** Nucleus (X)

## Manual Assessments

Score	Original propositions		Synonym-derived propositions		Troponym-derived propositions	
	Grader A	Grader B	Grader A	Grader B	Grader A	Grader B
Correct	141	123	32	68	21	53
Incorrect	12	32	93	88	104	114
Ambiguous	33	13	16	0	22	0
Structural violation	64	82	0	0	0	1

There are 4 possible scorings for each proposition:  
 1 (correct, e.g. “Photosynthesis - needs - Carbon dioxide”)  
 X (incorrect, e.g. “DNA - transcribes – RNA”)  
 0 (ambiguous, e.g. “atom - is made of – neutron”)  
 S (structural violation, e.g. “ocans - evaporation – atmosphere”).

Structural violations were for grammar problems such as spelling errors (“ocans” instead of “oceans”) and linking phrases that do not contain a verb. Ambiguous scores were for propositions that required a plausible context of surrounding propositions in order to be correct.

## Polysemy and Correctness

Polysemy, definition: The number of different contexts a word can be used in.

Linking word polysemy counts greater than 20 correspond to proposition validity by a factor of 2.4. Low polysemy counts less than 5 correspond to a 10% rise in incorrectness.

## Data Analysis Objectives

Evaluate current performance of tool for assessing validity of concept map propositions (Robograder™).

Measure impact of using WordNet’s ® synonym sets to “amplify” Robograder™.

Assess validity of original and randomly generated linking phrase substitutions.

Measure association of polysemy (multiple common meanings of usage) with validity.

## Robograder™ Design & Performance

Instructors design a grading rubric (which can be continually re-edited & updated) in a client-side spreadsheet environment like Microsoft Excel®.

Rubrics are automatically synchronized with Robograder™.

26% (9211 of 35404) of the propositions on the C-TOOLS server were gradable by Robograder™.

Robograder™ treats WordNet® synonyms as “equally valid”; this amounts to 10% (971 of the 9211) more propositions that are graded by Robograder™. We predict this does not introduce false gradings since students were observed selecting similar word senses.

## Correctness Among Synsets

The distribution of correctness across multiple synsets for the same word was not significantly different from arbitrarily chosen words (Grader A:  $\chi^2 = 1.59$ ; Grader B:  $\chi^2 = 2.07$ ).

### Loss of correctness inside synsets

Synonym derivatives:  
**15%** (grader A) and **14%** (grader B)  
 Troponym derivatives:  
**67%** (grader A) and **49%** (grader B)

### Conservation of correctness inside synsets

Synonym derivatives:  
**27%** (grader A) and **31%** (grader B)  
 Troponym derivatives:  
**100%** (grader A) and **81%** (grader B)

## Summary

Correctness is random across different synsets and conserved inside synsets.

Valid propositions tend to use common words.

Future: C-TOOLS is an open source tool to develop and refine measures for automated assessment.

