Visualisation of Verb Dependencies
in the EAP Classroom

Yulia Svetashova & Heike da Silva Cardoso

Uni Tübingen

June 21, 2014
Motivation

How to Use Concur in a Sentence

By Aya Hadano, eHow Contributor

Attorneys often use the word "concur" in legal documents. It has a long history and its use has evolved over the years. Since "concur" is rarely used in daily conversations, it is important to understand how to correctly use it in a sentence. According to the dictionary, the most common definition of "concur" is "to accord in opinion" or "agree." The most proper way to use "concur" in a sentence is to replace the word "agree."

argue that or argue
argue over or argue about
argue on or argue about
argue with or argue against

Hello to all

I wonder if someone could explain to me how to choose? Is there a context, a nuance that can guide the choice?

For instance between these two sentences, is one better (or different) than the other?

A- Their comparison will contribute to highlight the various drivers of adaptive capacity

B- Their comparison will contribute to highlighting the various drivers of adaptive capacity

Thanks in advance

Fénice
Visualisation of Verb Dependencies
in the British Academic Written English Corpus (BAWE)

Select a comparison type:
- level of study
- disciplinary group

Select a verb:
- alphabetically
- by frequency

I. Processing chain: tokenisation, POS-tagging, parsing

II. Comparison

III. Frequency data

IV. Dependencies: types and dependents

V. Concordance

Social Sciences

accentuate (59)
Motivation

Introduction to Dependency Parsing and Verb Dependencies

BAWE Corpus

Visualisation

Usage in the EAP Classroom

Further Work
What is a Dependency Grammar?

In a dependency grammar the verb is seen as the governor or head of a sentence.

Bell, based in Los Angeles, makes and distributes electronic, computer and building products [De Marneffe and Manning (2008)]
Verb Dependencies

Every verb needs a certain number and certain types of arguments, for example:
to eat 2 arguments of which one is optional

- Paul is eating dinner.
- Paul is eating.

while to devour needs two arguments:
- Paul devoured his steak.

and to give needs 3:
- Paul gave Mary the book.
Importance of Arguments

Arguments are essential to the understanding of a verb, because

1. they show us how it can be used and in which context and
2. how the meaning of a verb changes when different arguments are used
3. word sense disambiguation
Why are Verb Dependencies Important in the EAP Classroom?

- How a verb is used helps us understand it better.
- If we see how verb usage changes from level 1 to level 4 it helps us deduce what students have to learn and what we can expect of a writer.
- By comparing verb usage between different disciplines we can see variations.
The Stanford Parser [De Marneffe and Manning(2008)] is an algorithm, designed at Stanford University, CA. It takes a sentence as input and gives back a sentence which is annotated with Part of Speech tags \(^1\) and the dependency structure of a sentence.

\(^1\)POS tags denote the grammatical category of a word e.g. eat = verb, Paul = noun, etc.
Dependency Types

There are different types of dependency relations to a verb, like subject, direct object, indirect object etc.

The Stanford parser assigns these dependency relations to all words in the sentence.

Like in the following sentence:

’She gave me a book’

iobj\(^2\) (gave, me) dobj\(^3\) (gave, book) nsubj\(^4\) (gave, She)

\(^2\)indirect object
\(^3\)direct object
\(^4\)nominal subject
The Stanford Dependency Parser uses a set of approximately 50 grammatical relations.

- These relations are strictly binary.
- They hold between a governor and a dependent.

We chose a subset dependency types, which denote relations between verbs and their arguments.
We used the BAWE corpus [Nesi and Gardner(2012)] of EAP and ran the Stanford Parser for Dependencies on it. This gave us an annotated corpus including POS tags and dependency relations. This annotated corpus was used for queries to be able to filter verbs and their dependents for different disciplines and levels of writing.
## BAWE Corpus

<table>
<thead>
<tr>
<th>Disciplinary Group</th>
<th>Yr 1</th>
<th>Yr 2</th>
<th>Yr 3</th>
<th>Masters</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arts and Humanities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students</td>
<td>101</td>
<td>83</td>
<td>61</td>
<td>23</td>
<td>268</td>
</tr>
<tr>
<td>Assignments</td>
<td>239</td>
<td>228</td>
<td>160</td>
<td>78</td>
<td>705</td>
</tr>
<tr>
<td>Texts</td>
<td>259</td>
<td>231</td>
<td>161</td>
<td>83</td>
<td>734</td>
</tr>
<tr>
<td>Words</td>
<td>468,353</td>
<td>583,617</td>
<td>427,942</td>
<td>234,206</td>
<td>1,714,118</td>
</tr>
<tr>
<td><strong>Life Sciences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students</td>
<td>74</td>
<td>71</td>
<td>42</td>
<td>46</td>
<td>233</td>
</tr>
<tr>
<td>Assignments</td>
<td>180</td>
<td>193</td>
<td>113</td>
<td>197</td>
<td>683</td>
</tr>
<tr>
<td>Texts</td>
<td>191</td>
<td>208</td>
<td>119</td>
<td>203</td>
<td>721</td>
</tr>
<tr>
<td>Words</td>
<td>299,370</td>
<td>408,070</td>
<td>263,668</td>
<td>441,283</td>
<td>1,412,391</td>
</tr>
<tr>
<td><strong>Physical Sciences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students</td>
<td>73</td>
<td>60</td>
<td>56</td>
<td>36</td>
<td>225</td>
</tr>
<tr>
<td>Assignments</td>
<td>181</td>
<td>149</td>
<td>156</td>
<td>110</td>
<td>596</td>
</tr>
<tr>
<td>Texts</td>
<td>186</td>
<td>156</td>
<td>169</td>
<td>129</td>
<td>640</td>
</tr>
<tr>
<td>Words</td>
<td>300,989</td>
<td>314,331</td>
<td>426,431</td>
<td>339,605</td>
<td>1,381,356</td>
</tr>
<tr>
<td><strong>Social Sciences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students</td>
<td>85</td>
<td>88</td>
<td>75</td>
<td>62</td>
<td>313</td>
</tr>
<tr>
<td>Assignments</td>
<td>207</td>
<td>197</td>
<td>162</td>
<td>202</td>
<td>777</td>
</tr>
<tr>
<td>Texts</td>
<td>218</td>
<td>202</td>
<td>169</td>
<td>204</td>
<td>802</td>
</tr>
<tr>
<td>Words</td>
<td>371,473</td>
<td>475,668</td>
<td>440,674</td>
<td>688,921</td>
<td>1,999,130</td>
</tr>
<tr>
<td><strong>Total Students</strong></td>
<td>333</td>
<td>302</td>
<td>234</td>
<td>167</td>
<td>1039</td>
</tr>
<tr>
<td><strong>Total Assignments</strong></td>
<td>807</td>
<td>767</td>
<td>591</td>
<td>6587</td>
<td>2761</td>
</tr>
<tr>
<td><strong>Total Texts</strong></td>
<td>854</td>
<td>797</td>
<td>618</td>
<td>619</td>
<td>2897</td>
</tr>
<tr>
<td><strong>Total Words</strong></td>
<td>1,440,185</td>
<td>1,781,686</td>
<td>1,558,715</td>
<td>1,704,015</td>
<td>6,506,995</td>
</tr>
</tbody>
</table>
Processing

BAWE Corpus
Consisting of
2761 essays
in TEI XML format
Divided into
4 disciplines
&
4 levels of writing

Stanford Core NLP
For
POS Tagging
And
Dependency Parse

Annotated Corpus
XML

Java Program
To create a
Single XML Corpus
Including
only the
relevant information

Queries

List of
Verbs
&
their
Dependents
Why Visualise

1. to visualise patterns we already know exist
2. to visualise structures and detect patterns
This is a comparative visualisation

1. it shows differences in verb usage
2. it helps to identify specific patterns
Comparison Types

The user can choose between

- level of study comparison: 1, 2, 3, 4 (UG, PG)
- disciplinary groups comparison: Arts and Humanities, Social Sciences, Life Sciences and Medicine, Physical Sciences
Lists of Verbs

The user is provided with a list of all verbs in the corpus sorted

- by frequency
- alphabetically

Verb List Sorted by Frequency
Corpus Statistics: Disciplinary Group

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arts and Humanities</td>
<td>288,503</td>
</tr>
<tr>
<td>Life Sciences</td>
<td>238,205</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>337,341</td>
</tr>
<tr>
<td>Physical Sciences</td>
<td>235,215</td>
</tr>
<tr>
<td><strong>Total (tagged as) verbs:</strong></td>
<td><strong>1,099,264</strong></td>
</tr>
<tr>
<td><strong>Total number of words:</strong></td>
<td><strong>6,506,995</strong></td>
</tr>
</tbody>
</table>

\[ \mu = 274,816 \]

\[ \approx 17\% \]
## Corpus Statistics: Level of Study

<table>
<thead>
<tr>
<th>Level of Study</th>
<th>Number of Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>252,560</td>
</tr>
<tr>
<td>2</td>
<td>308,791</td>
</tr>
<tr>
<td>3</td>
<td>266,834</td>
</tr>
<tr>
<td>4</td>
<td>267,364</td>
</tr>
<tr>
<td>unknown</td>
<td>3,715</td>
</tr>
</tbody>
</table>

Total (tagged as) verbs: 1,099,264
Total number of words: 6,506,995

\[ \mu = 273,887.25 \]

\[ \approx 17\% \]
Information Encoded in the Dependency Tree

- Verb lemma, its frequency, POS tags, and forms
  - (agree, level 1) frequency: 256;
  - POS tags: VB, VBN, VBP, VBG, VBD, VBZ;
  - tokens: Agreed, agree, agreed, agreeing, agrees

- Dependency relation types, their frequencies
  - (agree, level 1) type: nominal subject; frequency: 198;
  - type: clausal complement; frequency: 95, etc.

- Dependents’ tokens, POS tags, lemmas, their frequencies
- Contexts of usage
Verb Dependencies Tree

Verb Dependents: Collapsible Tree and KWIC
Verb Dependencies Tree

Arts and Humanities

Prepositional Modifiers: Tree and KWIC

For example, they **object** that if the person is not identical to the human animal, then who is sitting in a chair typing these words, the animal or the person?

Tim Crane **objected** to Putnam's Twin Earth theory.

It may be **objected** by proponents of the wager argument that the existence of a made up God is far less probable than the God of Christian theology, but as long as the rewards of worshipping this made up God are infinite, then the value of its outcome is infinite, the same value which worshipping the Christian God carries; If we were to
Design Decisions

- Visual variable: **color**
- **selective** interpretation task:
  verbs in the KWIC part are colored red to cause "pop-up" effect
Visualisation of Verb Dependencies

Design Decisions

- Visual variable: **color**
- **associative**

interpretation task: the dependency relation types are grouped by meaning (e.g., the group of object tags circles are colored orange, subject group, - yellow, etc.)
Design Decisions

- Visual variable: **size**
- **selective** visual interpretation task: elements with higher frequencies have a bigger circle radius
Verb Dependencies Tree

- Hovering the node circle → additional information
- Levels of tree hierarchy: verb lemma, relation type, dependent

![Verb Dependencies Tree Diagram]

Additional Information
Interaction Types (Yi et al., 2007)

- Select: the verb is selected from a verb list
- Explore: the dependencies information is returned in a form of a tree
- Reconfigure: nodes are collapsible, the user can focus on the most interesting piece of data
- Abstract (Elaborate): data is shown in the tooltips
- Filter: filtering by frequency and alphabetically is possible
Comparison: Disciplinary Group

Dependencies Comparison
Comparison: Disciplinary Group

Social Sciences

Glorify

Dobj

AUX

To

Importance

Physical Sciences

Arts and Humanities

Glorify

Dobj

Nsubj

To

Will

Was

Herds

State

Everything

Framework

Himself

Thru

War

Highlands

Weakness

That

Sacrifice

Significance

Names

Name

Will

Experience

Delineation

Schopenhauer

Warner

Memorials

Benedict

Hot

Goal

Army

Who

Life Sciences
Comparison: Level

Dependencies Comparison
Usage

Two example usages would be to identify

1. styles of writing
2. error patterns

The different types of comparison help identify these features in different fields.
Usage in the EAP Classroom

1. by teachers to help understand the level of writing of their students
2. to see what can be expected in a specific discipline or a level of writing
3. detect frequent errors
4. by students to understand possible word usages
Further Work

- choice options
- more flexible data manipulation
- zooming/scaling
- query data export
- multiple verbs comparison
- adding features to be encoded by visual variables
- pipeline tool
- new levels in tree hierarchy (POS)
Summary

We think that

- Visualisation can be very useful to help identify patterns
- Verb dependency structures capture an essential part of the sentence structure

Therefore we believe that the visualisation of verb dependency structures is of interest in many fields
Further Work

Yulia Svetashova & Heike da Silva Cardoso
References

MST Carpendale.
Considering visual variables as a basis for information visualisation.
*Computer Science TR# 2001-693, 16, 2003.*

Marie-Catherine De Marneffe and Christopher D Manning.
Stanford typed dependencies manual.

Hilary Nesi and Sheena Gardner.
*Genres across the disciplines: Student writing in higher education.*

Ji Soo Yi, Youn ah Kang, John T Stasko, and Julie A Jacko.
Toward a deeper understanding of the role of interaction in information visualization.