Changes in thematic choice with developing EFL proficiency

Mick O’Donnell
Universidad Autónoma de Madrid

After the revolution, good hamburgers were ...
Aims

- We are a group of University English teachers in Spain
- Studying learner writing to understand how learners learn
- And thus how to improve the learning environment for our learners
Our Learner Corpus

- We make use of two corpora of short essays written by learners of English in Spanish Universities
  - The WriCLE corpus (UAM): 500,000 words by students in English degree. (Rollinson and Mendikoetxea 2008)
  - The UPV Learner Corpus (UPV): 150,000 words of shorter texts by ESP students (Andreu Andrés et al. 2010)
- All texts associated with proficiency level using Oxford Quick Placement Test
Analysis

We study the learner writing through two methods of analysis:

- (Automatic) **Syntactic Analysis** (Mood, Theme, Transitivity) via UAM CorpusTool
  - To see what learners do/don’t do.

- (Manual) **Error Analysis**: via 8 human coders
  - To see what learners are struggling with
Our Prior Studies

- Errors (MacDonald-Murcia, etc.)
- Modality (Garcia 2011)
- Transitivity (O’Donnell 2012)
- Lexical Errors (Mediero, 2013)
- Tense-Aspect (O’Donnell 2013)
- Article Use (Dotti 2014)
- Theme (O’Donnell 2014)
Our Prior Studies

- **Errors** (MacDonald-Murcia, etc.)
- **Modality** (Garcia 2011)
- **Transitivity** (O’Donnell 2012)
- **Lexical Errors** (Mediero, 2013)
- **Tense-Aspect** (O’Donnell 2013)
- **Article Use** (Dotti 2014)
- **Theme** (O’Donnell 2014)

This talk addresses this study
2. Linguistic Model of Theme

After the revolution, life was not the same
Linguistic Model

I assume Halliday’s model, as presented in “Introduction to Functional Grammar”, 4th Edition (Halliday and Matthiessen).

In declaratives, Theme includes all clausal elements up to and including the first experiential element (most typically the Subject).

So, Textual and Interpersonal elements may precede:

<table>
<thead>
<tr>
<th>Unfortunately</th>
<th>however</th>
<th>the revolution</th>
<th>failed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theme</td>
<td>Rheme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpers.</td>
<td>Textual</td>
<td>Topical</td>
<td></td>
</tr>
</tbody>
</table>
Linguistic Model
Topical choices

In declaratives:

- unmarked Topical theme is SUBJECT:
  - \( \text{John likes coffee in the morning.} \)

- fronted Adjunct:
  - \( \text{In the morning John likes coffee.} \)

- fronted Complement:
  - \( \text{Coffee, John likes.} \)

- fronted Dependent Clause:
  - \( \text{Because I drank too much coffee, I cannot sleep.} \)
Linguistic Model

Topical choices

In declaratives:

- unmarked Topical theme is SUBJECT:
  -> **John** likes coffee in the morning.

- fronted Adjunct:
  -> **In the morning** John likes coffee.

- fronted Complement:
  -> **Coffee, John** likes.

- fronted Dependent Clause:
  -> **Because I drank too much coffee, I cannot** sleep.
Linguistic Model

Topical choices

Coordinated clauses with Elliptical Subject:

\[ \rightarrow I \text{ drink coffee but usually hate the taste.} \]

- Taken to have no Topical Theme (the Topical theme of the first clause carries over to the coordinated clause.)
Linguistic Model

Topical choices

Not handled:

- Predicated themes (it-cleft):
  -> *It is* John *that likes coffee.*

- Thematic Equatives (wh-clefts):
  -> *What john likes* if coffee.
  -> *Coffee* is what John likes.

- Thematicised comment:
  -> *It is true* that John is good.
3. Automating Theme Analysis
Automating Theme Analysis

- UAMCT calls Stanford Parser to produce a basic syntactic analysis:
Automating Theme Analysis

- UAMCT translates Stanford analysis into something closer to SFL Mood Analysis:
Automating Theme Analysis

- UAMCT translates Mood analysis into Theme-Rheme analysis:

\[\text{On the other hand, the democrats were not strong either.}\]

<table>
<thead>
<tr>
<th>Theme</th>
<th>Rheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Textual</td>
<td>Topical</td>
</tr>
</tbody>
</table>
Automating Theme Analysis: 

Procedure

- For each constituent at the front of the clause:
  - If **Subject**, code as **Topical**
  - If **Adjunct,**
    - If on list of Textual markers, or grammatically a conjunction, code as **Textual**
    - if on list of Mood markers, or grammatically modal, code as **Interpersonal**
Automating Theme Analysis

Currently:

- 144 Textual markers
- 126 Interpersonal markers
- User can add their own to the list.

<table>
<thead>
<tr>
<th>textual</th>
<th>addition</th>
<th>and</th>
<th>additionally</th>
<th>also</th>
<th>moreover</th>
<th>in addition</th>
<th>besides</th>
<th>as well</th>
</tr>
</thead>
<tbody>
<tr>
<td>textual</td>
<td>consequence</td>
<td>as a consequence</td>
<td>as a result of</td>
<td>because of that</td>
<td>because of</td>
<td>due to this</td>
<td>due to this</td>
<td>for that</td>
</tr>
<tr>
<td>textual</td>
<td>summative</td>
<td>in short</td>
<td>briefly</td>
<td>to sum up</td>
<td>in summary</td>
<td>to summarize</td>
<td>to summarise</td>
<td></td>
</tr>
<tr>
<td>textual</td>
<td>appositive</td>
<td>for example</td>
<td>as an</td>
<td>that is</td>
<td>in other</td>
<td>for instance</td>
<td>as I said before</td>
<td></td>
</tr>
<tr>
<td>textual</td>
<td>corrective</td>
<td>rather</td>
<td>or rather</td>
<td>to be precise</td>
<td>at least</td>
<td>strictly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>textual</td>
<td>adversative</td>
<td>but</td>
<td>however</td>
<td>on the other</td>
<td>conversely</td>
<td>by the other</td>
<td>in the other</td>
<td>on the</td>
</tr>
<tr>
<td>textual</td>
<td>contextualising</td>
<td>as to that</td>
<td>in this</td>
<td>here</td>
<td>in that way</td>
<td>in this way</td>
<td>in this way</td>
<td>in this</td>
</tr>
<tr>
<td>textual</td>
<td>sequencing</td>
<td>firstly</td>
<td>secondly</td>
<td>thirdly</td>
<td>fourthly</td>
<td>fifthly</td>
<td>sixthly</td>
<td>forthly</td>
</tr>
<tr>
<td>textual</td>
<td>comparing</td>
<td>in the one hand</td>
<td>on one hand</td>
<td>on the one hand</td>
<td>in one hand</td>
<td>for one hand</td>
<td>in the one hand</td>
<td></td>
</tr>
<tr>
<td>textual</td>
<td>concluding</td>
<td>finally</td>
<td>in conclusion</td>
<td>as a conclusion</td>
<td>to conclude</td>
<td>finally</td>
<td>last</td>
<td>last but not</td>
</tr>
<tr>
<td>textual</td>
<td>concessive</td>
<td>yet</td>
<td>despite this</td>
<td>despite that</td>
<td>nevertheless</td>
<td>in spite of</td>
<td>in spite of</td>
<td>nonetheless</td>
</tr>
<tr>
<td>textual</td>
<td>dismissive</td>
<td>in any case</td>
<td>anyway</td>
<td>leaving that</td>
<td>apart from this</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>textual</td>
<td>actuality</td>
<td>actually</td>
<td>in fact</td>
<td>indeed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>textual</td>
<td>likewise</td>
<td>likewise</td>
<td>in the same way</td>
<td>in the same vein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>textual</td>
<td>alternatively</td>
<td>instead</td>
<td>alternatively</td>
<td>or</td>
<td>otherwise</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
On the other hand, the Democrats were not strong either.
Automating Theme Analysis

- Mapping Mood to Theme:
  - Left with T-unit, Textual, Topical and Interpersonal segments.
  - Their features are changed to reflect their thematic nature (mood features dropped)
  - All other Mood structure constituents dropped
  - New Theme-group constituent created to group Textual, Interpersonal, Topical.
  - Rest of clause tagged as Rheme.
In Madrid about 200 people died in a terrorist attack caused by the participation of Spain in the war of Irak.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Rheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topical</td>
<td></td>
</tr>
</tbody>
</table>

London was also affected by this kind of actions, but fortunately no deaths had to be regreted.

<table>
<thead>
<tr>
<th>Element</th>
<th>Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theme</td>
<td>Theme</td>
</tr>
<tr>
<td>Rheme</td>
<td>Rheme</td>
</tr>
<tr>
<td>Textual</td>
<td>Interpersonal</td>
</tr>
<tr>
<td>Topical</td>
<td></td>
</tr>
</tbody>
</table>

This is the most extended position amongst the world population, maybe because of these lamentable incidents.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Rheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topical</td>
<td></td>
</tr>
</tbody>
</table>
Automating Theme Analysis

Handling Learner variation

- Stanford parser actually quite tolerant of learner errors in most cases

- Textual/Interpersonal marker dictionary enriched with learner variants:

<table>
<thead>
<tr>
<th>on the other hand</th>
<th>on other hand</th>
<th>in the second hand</th>
</tr>
</thead>
<tbody>
<tr>
<td>by the other hand</td>
<td>by other hand</td>
<td>for another hand</td>
</tr>
<tr>
<td>in the other hand</td>
<td>one the other hand</td>
<td>by other hand</td>
</tr>
<tr>
<td>on the other side</td>
<td>to the other hand</td>
<td>in a other hand</td>
</tr>
<tr>
<td>in other hand</td>
<td>in another hand</td>
<td></td>
</tr>
</tbody>
</table>
4. Results
Results (i): Number of T-units

<table>
<thead>
<tr>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>905</td>
<td>5018</td>
<td>12492</td>
<td>9376</td>
<td>5119</td>
<td>923</td>
</tr>
</tbody>
</table>
Results: Types of Topical Themes
(as percent of T-units)
Results: Most Frequent Topical Themes (as percent of T-units)
Results: Use of Textual Themes
(as percent of T-units)

Textual Theme

A1 | A2 | B1 | B2 | C2 | C2
---|----|----|----|----|----
20.00% | 22.00% | 24.00% | 26.00% | 28.00% | 30.00% | 32.00% | 34.00% | 36.00% | 38.00%
Results: Use of Textual Themes
(as percent of T-units)

<table>
<thead>
<tr>
<th>Textual Theme</th>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>905</td>
<td>5018</td>
<td>12492</td>
<td>9376</td>
<td>5119</td>
<td>923</td>
</tr>
</tbody>
</table>
Results: Common Textual markers 
(as percent of Textual elements)
Textual markers by semantics
Results: Textual markers by semantics
(as percent of T-units)

extending
arguing
structuring
Results: Textual markers by semantics (as percent of T-units)

Extending Textual Markers
Results: Textual markers by semantics
(as percent of T-units)

Arguing Textual Markers

- consequence
- corrective
- adversative
- concessive
- concessive
- alternatively

[Graph showing the percentages of different types of textual markers over different samples (A1 to C2).]
Results: Use of Interpersonal Themes (as percent of T-units)
Interpersonal markers by semantics

- probability
- usuality
- typicality
- obviousness
- expectation
- claim
- belief
  - without doubt
- opinion
  - in my opinion
- evidence
- factuality
- desirability
- sensibleness
- validity
- reaction
  - happily, sadly
- admission
  - frankly
- approximation
  - roughly, approximately
- tentativeness
Interpersonal Themes
(per 1000 T-units)

Frequent Interpersonal Themes

Instances per 1000

- in my opinion
- to me
- sometimes
- from my point of view
- of course
- maybe
- 's
- appears
- personally speaking
- in my view

A1 A2 B1 B2 C2 C2
Interpersonal Themes by Type
(as % of T-units)

![Graph showing interpersonal themes by type.](image-url)
Interpersonal Themes by Type
(as % of interpersonal tokens)
5. Discussion
Discussion

- Results not as clear-cut as I expected
- graphs show U-shapes, not clear rise or fall.
- Similar studies on Transitivity, tense-aspect, errors with corpus show clearer trends.
- Development of thematic proficiency obviously more complex than I thought.
Discussion

Topical Selection:

- No use of Complement themes
- Slight increase in Adjunct themes after initial slump
Discussion

Textual Themes:

❖ Fall in use of Textual markers with proficiency

❖ Mainly in the area of **Extending** markers ("and", "additionally").

❖ Increased use of **Argumentative** markers (adversative, concession, consequence)

❖ Some rise in **Structuring** markers ("secondly", etc.)
Discussion

Interpersonal Themes:

- In general, interpersonal themes more frequent with increasing proficiency.
- Evidentiality markers (probability etc.) rises then falls,
- Evaluative markers (honestly, sadly, etc.) fairly constant but not common.