A corpus-based grammatical analysis of modality in the writing of Spanish university students of English

Rebeca García González

Financiado por el Ministerio de Ciencia e Innovación (FFI2009-14436/FILO)
Introduction

Aim of paper:

• To explore how Spanish university students grammatically express modality in their writing across proficiency levels
Outline

• Research hypotheses
• Theoretical framework
• Corpus and procedures
• Findings
• Conclusions
Research Hypotheses

As students’ level of proficiency rises, I expect:

1. A higher quantity of modal markers
2. A wider variety of modal markers
3. A movement away from *verbal* to *non-verbal* modal means
Theoretical framework

A typological account of modality:

**Mood** vs. **Modal systems**

*Indicative* – Modal *A set of modal forms*

*Subjunctive* – Non-modal “Prototypical” means: Modal auxiliary verbs

(i.e. Romance languages) (i.e. English)

(Palmer, F., 2001; 2003)
Halliday’s views on modality:

<table>
<thead>
<tr>
<th>COMMODITY EXCHANGED</th>
<th>SPEECH FUNCTION</th>
<th>TYPE OF INTERMEDIACY</th>
<th>TYPICAL REALIZATION</th>
<th>EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>Proposition</td>
<td>Probability</td>
<td>-Finite Modal Operator</td>
<td>-They must have known</td>
</tr>
<tr>
<td></td>
<td>Question</td>
<td></td>
<td>-Modal Adjunct</td>
<td>-They certainly knew</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Both the above combined</td>
<td>-They certainly must have known</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goods and Services</td>
<td>Proposal</td>
<td>Frequency</td>
<td>-Finite Modal Operator</td>
<td>-It must happen</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Modal Adjunct</td>
<td>-It always happens</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Both the above combined</td>
<td>-It must always happen</td>
</tr>
<tr>
<td></td>
<td>Command</td>
<td>Obligation</td>
<td>-Finite Modal Operator</td>
<td>-You must be patient!</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Passive Verb Predicator</td>
<td>-You’re required to be patient!</td>
</tr>
<tr>
<td></td>
<td>Offer</td>
<td>Inclination</td>
<td>-Finite Modal Operator</td>
<td>-I must win!</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Adjective Predicator</td>
<td>-I’m determined to win!</td>
</tr>
</tbody>
</table>

(from Halliday and Matthiessen, 2004)
Nuyts’ model of modality:

- Cross-linguistic analysis of modality: German, Dutch and English
- Grammatical elements expressing modality:
  - Modal adverbs
  - Modal adjectives
  - Modal auxiliaries
  - Mental state predicates:
    - Descriptive vs Performative
    - Qualificational vs. Non-qualificational

(from Nuyts, 2000)
Corpus

- The **WriCLE** corpus: a *learner corpus* of Spanish university students (Rollinson and Mendikoetxea, 2010)
- A section of **458 essays** analyzed: 445,776 words (263 first-year-student essays; 195 third-year-student essays)
- Metadata included: **Learner Profile**, with the students’ level of proficiency – Oxford Quick Placement Test, UCLES(2001)
• Writers’ levels of proficiency: from A2 to C2, following the CEFRL (Council of Europe, 2001)
Mean clauses per essay across proficiency level

<table>
<thead>
<tr>
<th>Level</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>96.21</td>
<td>121.33</td>
<td>135.28</td>
<td>166.5</td>
<td>144.5</td>
</tr>
</tbody>
</table>

Mean clauses per essay across proficiency level
Procedures:

- **UAM CorpusTool** allows manual and automatic annotation of collections of text: pattern matching (O’Donnell, 2008)
  
  the software automatically assigns grammatical categories to modal markers following my taxonomy
Rule: \textbf{select Modality intention if Modality.\textit{lexical-verb} containing \textit{think}}
To understand more or less what children think, we have to...
Taxonomy of modal markers used:

- **modal-marker**
  - verbal-modality
    - modal-auxilliary
      - lexico-modal-auxilliary
    - lexical-verb
    - modal-idiom
  - non-verbal-modality
    - adjectival-phrase
    - adverbials
  - epistemic
    - likelihood
      - certainty
      - probability
      - possibility
    - obligation
      - inescapable-obligation
      - unfulfilled-obligation/advisability
    - necessity/lack-of-necessity
    - permission
  - deontic
    - ability
    - willingness
    - intention
  - dynamic
Verbal modal elements:

- **modal-auxilliary**
  - CAN, COULD, MAY, MIGHT, WILL, WOULD, SHALL, SHOULD, MUST and OUGHT TO

- **lexico-modal-auxilliary**
  - HAVE (GOT) TO, BE GOING TO, BE SUPPOSED TO, BE OBLIGED TO, BE REQUIRED TO, BE BOUND TO, BE ALLOWED TO and NEED

- **lexical-verb**
  - I THINK, I BELIEVE, I SUPPOSE, I GUESS, I FEEL, I FIND, I EXPECT, I KNOW, I RECKON, I CONSIDER, I INTEND and I DOUBT

- **modal-idiom**
  - HAD BETTER, WOULD RATHER and WOULD SOONER

(from Greenbaum et al., 1985)
Non-verbal modal elements:

**adjectival-phrase**

ABLE, UNABLE, TRUE,
FALSE, NECESSARY,
UNNECESSARY, POSSIBLE,
IMPOSSIBLE, PROBABLE,
IMPROBABLE, CLEAR, LIKELY,
UNLIKELY, SURE, UNSURE,
CERTAIN, UNCERTAIN, PERMISSIBLE,
WILLING, UNWILLING, DETERMINED,
IMPERATIVE, BOUND

**adverbials**

MAYBE, PERHAPS, POSSIBLY,
DEFINITELY, CERTAINLY, SUPPOSEDLY,
SURELY, UNDOUBTEDLY, DOUBTLESS,
DOUBTLESSLY, LIKELY, PRESUMABLY,
ARGUABLY, CONCEIVABLY, INDEED,
FOR SURE, FOR CERTAIN, OF COURSE,
WITHOUT DOUBT, PROBABLY and NECESSARILY.

(from Greenbaum et al., 1985)
Modal tokens per clause

A2: 0.20
B1: 0.1866
B2: 0.1895
C1: 0.1698
C2: 0.1757
Hypothesis 1

- As students’ level of proficiency rises, they will use a higher quantity of modal markers – PROVEN WRONG
Hypothesis 3:

- As students’ level of proficiency rises, I expect a movement away from *verbal* to *non-verbal* modal means
Verbal vs Non-Verbal modality

- N: 10287 (92%)
- N: 873 (8%)

Verbal Modality
Non-Verbal Modality
These results show that hypothesis 3 is RIGHT
Verbal Modality per total modal markers used

- A2: 88.60%
- B1: 93.20%
- B2: 91.40%
- C1: 90.90%
- C2: 88.80%

Non-Verbal Modality per total modal markers used

- A2: 11.40%
- B1: 6.80%
- B2: 8.60%
- C1: 9.10%
- C2: 11.20%
Hypothesis 2:

- As students’ level of proficiency rises, I expect them to use a wider variety of modal markers.
### Modal aux per total modals used

<table>
<thead>
<tr>
<th>Level</th>
<th>Modal auxiliaries per total modal markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>68,22%</td>
</tr>
<tr>
<td>B1</td>
<td>71,92%</td>
</tr>
<tr>
<td>B2</td>
<td>68,56%</td>
</tr>
<tr>
<td>C1</td>
<td>71,67%</td>
</tr>
<tr>
<td>C2</td>
<td>75,87%</td>
</tr>
</tbody>
</table>

### Lexico-modals per total modal markers

<table>
<thead>
<tr>
<th>Level</th>
<th>Lexico-modal per total modal markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>9,34%</td>
</tr>
<tr>
<td>B1</td>
<td>9,47%</td>
</tr>
<tr>
<td>B2</td>
<td>9,68%</td>
</tr>
<tr>
<td>C1</td>
<td>7,93%</td>
</tr>
<tr>
<td>C2</td>
<td>6,54%</td>
</tr>
</tbody>
</table>

### Lexical Vbs per modal markers

<table>
<thead>
<tr>
<th>Level</th>
<th>Lexical Verbs per total modal markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>10,99%</td>
</tr>
<tr>
<td>B1</td>
<td>11,85%</td>
</tr>
<tr>
<td>B2</td>
<td>13,19%</td>
</tr>
<tr>
<td>C1</td>
<td>12,24%</td>
</tr>
<tr>
<td>C2</td>
<td>6,34%</td>
</tr>
</tbody>
</table>
• There are no clear patterns, because all the grammatical categories established as modal are used to some degree at all levels.

• What does change is that the more advanced levels, especially B2 and C1, use a wider range of tokens within each of these categories
Types of lexico-modal verbs used by C1 students

- have to: 52.59%
- be going to: 20.73%
- need to: 17.04%
- be supposed to: 5.18%
- be obliged to: 1.48%
- be bound to: 1.48%
- be required to: 1.48%

Types of lexico-modal verbs used by C2 students

- have to: 74.99%
- be going to: 15.61%
- need to: 9.37%
Types of adverbials used by C1 students

Types of adverbials used by C2 students
Hypothesis 2:

- As students’ level of proficiency rises, I expect them to use a wider variety of modal markers:

  Increasing use of non-verbal modality and more variety of tokens within each category: it proves RIGHT.
Conclusions

As students’s proficiency level rises,

• A fall in the use of total modal markers:

- **Verbal modal elements: decreasing** presence
  ***BUT their number is the largest across proficiency levels***

- **Non-verbal modal elements: increasing** presence
• Proportion of modal markers per clause may be diluted in the advanced levels (they write more)

• A movement towards categories not included in my taxonomy? (i.e. evidentiality, frequency, reporting verbs, mental verbs different from the ones included)

• More quantity does not mean more quality, e.g. decreasing use of be going to
• The tokens used in each grammatical class show more grammatical and pragmatic accuracy in the advanced levels, e.g. low-proficient students overuse *can* and *will* at the expense of *would* and *should* (*modal aux.*); decreasing use of *maybe* vs. increasing use of *probably* (*adverbials*)

  > notions of *register* and *genre*

• A wider range of tokens within each grammatical class as level of proficiency rises (except for C2 students)
Thank you!

Rebeca García González
rebeca.garciagonzalez@estudiante.uam.es