A Corpus-based Comparison of Albanian and Italian Student Academic Writings in L1 and English as L2: Introducing Hedges and Boosters within Modalization, Modulation and Degree

Vincenzo Dheskali
Sixth Semester PhD Student
Chemnitz University of Technology
vincenzo.dheskali@s2015.tu-chemnitz.de
Hedges and boosters: My working definition

Hedges (e.g. possibly, almost, I think) and boosters (e.g. certainly, completely, demonstrate) are numerous lexical and non-lexical devices which express different degrees of authors’ direct and indirect commitment regarding the probability and usuality of the expressed proposition. They are modalization, modulation and degree devices that interweave interpersonal and ideational socio-semiotic processes on a semantic level (approximators), pragmatic level (shields), and display an interaction of both. They express different forms of manifestation, orientation, prosody of ‘modality’, syntactic positioning, approximation, shields and polarity across and within different cultural and linguistic contexts of student academic writings.

I notice something which polemicists seemed to have missed: the demonstrable, elementary fact that Wilde is almost always right. (CIAOE15FBL_82)


stronger colour=clearer hedge and booster
weaker colour=less hedge and booster
Hedges and boosters: My working definition

Hedges (e.g. possibly, almost, I think) and boosters (e.g. certainly, completely, demonstrate) are numerous lexical and non-lexical devices which express different degrees of authors’ direct and indirect commitment regarding the probability and usuality of the expressed proposition. They are modalization, modulation and degree devices that interweave interpersonal and ideational socio-semiotic processes on a semantic level (approximators), pragmatic level (shields), and display an interaction of both. They express different forms of manifestation, orientation, prosody of ‘modality’, syntactic positioning, approximation, shields and polarity across and within cultural and linguistic contexts of student academic writings.

I notice something which polemicists seemed to have missed: the demonstrable, elementary fact that Wilde is almost always right. (CIAOE15FBL_82)

Methodology

- Corpus analysis

- Qualitative and quantitative comparison with the concordance software AntConc (Anothony 2014)
  - 15,000 hedging and boosting instances: copied to Excel and analysed based on Bondi (2008), Halliday and Matthiessen (2014), Lafuente Millàn (2008), Prince et al. (1980) and Quirk et al. (1985) and 10 Albanian, English and Italian dictionaries. Statistical measurements through Logistic Regression (with Rbuk, R), Log-Likelihood and relative frequencies x 1M words.

- Four corpora consisting of L1 and L2 writings by Albanian and Italian students
Picture 1: A map with the geographical distribution of my Italian corpora, CIAO (ita) and CIAOE (eng)
### Table 1: Albanian and Italian corpora including respective word totals according to AntConc

<table>
<thead>
<tr>
<th>Type of Papers</th>
<th>Italian Corpus (CIAO)</th>
<th>Italian English Corpus (CIAOE)</th>
<th>Type of Papers</th>
<th>Italian Corpus (CIAO)</th>
<th>Italian English Corpus (CIAOE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># papers</td>
<td># words</td>
<td></td>
<td># papers</td>
<td># words</td>
</tr>
<tr>
<td>PhD theses</td>
<td>54</td>
<td>2,035,346</td>
<td>PhD theses</td>
<td>57</td>
<td>1,984,929</td>
</tr>
<tr>
<td>MA theses</td>
<td>28</td>
<td>884,886</td>
<td>MA theses</td>
<td>26</td>
<td>712,014</td>
</tr>
<tr>
<td>BA theses</td>
<td>2</td>
<td>42,209</td>
<td>MA term papers</td>
<td>3</td>
<td>7,744</td>
</tr>
<tr>
<td>BA term papers</td>
<td>3</td>
<td>6,958</td>
<td>BA theses</td>
<td>4</td>
<td>67,778</td>
</tr>
<tr>
<td>males</td>
<td>40</td>
<td>1,392,173</td>
<td>males</td>
<td>45</td>
<td>1,325,840</td>
</tr>
<tr>
<td>females</td>
<td>47</td>
<td>1,577,226</td>
<td>females</td>
<td>45</td>
<td>1,446,625</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of Papers</th>
<th>Albanian Corpus (CAR)</th>
<th>Albanian English Corpus (CARE)</th>
<th>Type of Papers</th>
<th>Albanian Corpus (CAR)</th>
<th>Albanian English Corpus (CARE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># papers</td>
<td># words</td>
<td></td>
<td># papers</td>
<td># words</td>
</tr>
<tr>
<td>PhD theses</td>
<td>52</td>
<td>2,285,498</td>
<td>PhD theses</td>
<td>9</td>
<td>337,184</td>
</tr>
<tr>
<td>MA theses</td>
<td>10</td>
<td>144,225</td>
<td>MA theses</td>
<td>10</td>
<td>144,225</td>
</tr>
<tr>
<td>MA term papers</td>
<td>8</td>
<td>22,214</td>
<td>MA term papers</td>
<td>8</td>
<td>22,214</td>
</tr>
<tr>
<td>BA theses</td>
<td>13</td>
<td>110,054</td>
<td>BA theses</td>
<td>13</td>
<td>110,054</td>
</tr>
<tr>
<td>BA term papers</td>
<td>1</td>
<td>2,620</td>
<td>BA term papers</td>
<td>1</td>
<td>2,620</td>
</tr>
<tr>
<td>males</td>
<td>26</td>
<td>1,108,837</td>
<td>males</td>
<td>28</td>
<td>485,807</td>
</tr>
<tr>
<td>females</td>
<td>26</td>
<td>1,176,661</td>
<td>females</td>
<td>13</td>
<td>130,490</td>
</tr>
</tbody>
</table>
### Table 2: Albanian and Italian corpora according to sections (genres), texts and words

<table>
<thead>
<tr>
<th>sections</th>
<th># papers</th>
<th># words</th>
<th>sections</th>
<th># papers</th>
<th># words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italian Corpus (CIAO)</td>
<td></td>
<td></td>
<td>Italian English Corpus (CIAOE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lang. &amp; Lit.</td>
<td>40</td>
<td>1,402,214</td>
<td>Lang. &amp; Lit.</td>
<td>41</td>
<td>1,197,841</td>
</tr>
<tr>
<td>Social Studies</td>
<td>8</td>
<td><strong>483,368</strong></td>
<td>Social Studies</td>
<td>8</td>
<td>462,368</td>
</tr>
<tr>
<td>Medicine</td>
<td>9</td>
<td><strong>86,084</strong></td>
<td>Medicine</td>
<td>10</td>
<td>116,077</td>
</tr>
<tr>
<td>Chemistry</td>
<td>8</td>
<td>233,434</td>
<td>Chemistry</td>
<td>8</td>
<td>199,209</td>
</tr>
<tr>
<td>Physics</td>
<td>6</td>
<td>172,326</td>
<td>Physics</td>
<td>8</td>
<td>257,065</td>
</tr>
<tr>
<td>Economics</td>
<td>8</td>
<td>386,148</td>
<td>Economics</td>
<td>8</td>
<td>267,383</td>
</tr>
<tr>
<td>Math. &amp; Inform.</td>
<td>8</td>
<td>205,825</td>
<td>Math. &amp; Inform.</td>
<td>7</td>
<td>264,778</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td><strong>87</strong></td>
<td><strong>2,969,399</strong></td>
<td><strong>total</strong></td>
<td><strong>90</strong></td>
<td><strong>2,772,465</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sections</th>
<th># papers</th>
<th># words</th>
<th>sections</th>
<th># papers</th>
<th># words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albanian Corpus (CAR)</td>
<td></td>
<td></td>
<td>Albanian English Corpus (CARE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lang. &amp; Lit.</td>
<td>8</td>
<td>605,556</td>
<td>Lang. &amp; Lit.</td>
<td>33</td>
<td>293,899</td>
</tr>
<tr>
<td>Social Studies</td>
<td>8</td>
<td><strong>483,872</strong></td>
<td>Social Studies</td>
<td>3</td>
<td>127,883</td>
</tr>
<tr>
<td>Medicine (+M.Law)</td>
<td>3 (+1)</td>
<td><strong>103,037</strong></td>
<td>Chemistry</td>
<td>1</td>
<td><strong>69,513</strong></td>
</tr>
<tr>
<td>Chemistry</td>
<td>8</td>
<td>214,619</td>
<td>Physics</td>
<td>1</td>
<td><strong>14,979</strong></td>
</tr>
<tr>
<td>Biotechno. &amp; Physics</td>
<td>8</td>
<td>334,607</td>
<td>Economics</td>
<td>2</td>
<td>75,173</td>
</tr>
<tr>
<td>Economics</td>
<td>8</td>
<td>361,906</td>
<td>Informatics</td>
<td>1</td>
<td><strong>34,850</strong></td>
</tr>
<tr>
<td>Math. &amp; Inform.</td>
<td>8</td>
<td>181,901</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>total</strong></td>
<td><strong>52</strong></td>
<td><strong>2,285,498</strong></td>
<td><strong>total</strong></td>
<td><strong>41</strong></td>
<td><strong>616,297</strong></td>
</tr>
</tbody>
</table>

The table compares Albanian and Italian corpora across different sections (genres), with statistics on the number of papers and words for each category.
Key Concepts of my Study

- **Clause Complex**
  - (hypotactic/paratactic)
  - Mood System: Modality

- **Emphasizers**
  - Certainly
  - Completely, totally

- **Boosters**
  - Definitely, surely
  - High prob.

- **Hedges**
  - Medium/low prob.
  - Possibly, may

- **Shields**
  - Plausibility: I think
  - Attribution: according to

- **Approximators**
  - Degree: almost+adj.
  - Limitation: partly, almost+verb

- **Adaptors**
  - Quantity: around, almost+nr.

- **Rounders**
  - Frequency: sometimes, almost+time adv.

- **Degree Counter-expect.**

- **Orientation-manifestation**

- **Author Commitment**

- **Clustering/Prosody of Modality**

- **Towards Modalization**
  - Positive pole
  - Modality
  - Intensity
  - Negative pole

- **Proposals/Obligation**

- **Modulation**
  - Probability/usuality

- **Declarative**
  - Interrogative

- **We are working.**
  - Are you working?

(cf. Halliday and Matthiessen 2014; Lafuente 2008; Prince et al. 1982; Quirk et al. 1985; Bondi 2008: 39)
Key Concepts of my Study

- **positive pole**
  - modality
  - intensity

- **negative pole**
  - orientation
  - manifestation

- **boosters**
  - certainly
  - completely, totally
  - definitely, surely
  - high prob.
  - total deg.

- **hedges**
  - possibly, may
  - medium/low prob.
  - high/low deg.

- **shields**
  - I believe

- **approximators**
  - approximately

(cf. Halliday and Matthiessen 2014; Laffuente 2008; Prince et al. 1982; Quirk et al. 1985; Bondi 2008: 39)
Orientation (objective, subjective) and Manifestation (explicit, implicit)

1a). Subjective explicit: projecting mental clause, supported by an idea clause: *I guess, it is believed that, I don’t think* it was inappropriate for Federico, *I suppose* Lucrezia was disappointed by these events.

1b). Subjective implicit: commonly a finite (modal auxiliary) within the clause and mood: *Miri who could be the main actor, his friends may be influenced.*

2a). Objective implicit: clause, mood - usually carried by a mood adjunct, esp. a modal adverb: *certainly. / Placed in between the sentence e.g. in a neutral position: Trump will possibly lose the elections.*

2b). Objective explicit: usually expressed through a relational clause containing a factual carrier, clause etc.: *It is certain, it is probable that* (see Halliday and Matthiessen 2014: 688; Martin and White 2005: 130-131). / *Emphasized modal adverb e.g. placed initially: Obviously, it was successful., parenthetically.*

Values

<table>
<thead>
<tr>
<th>High:</th>
<th>Medium:</th>
<th>Low:</th>
</tr>
</thead>
<tbody>
<tr>
<td>certainly = <em>not possibly</em></td>
<td>probably = <em>not probably</em></td>
<td>possibly = <em>not certainly</em> (Halliday and Matthiessen 2014: 180).</td>
</tr>
</tbody>
</table>

*my theoretical additions*
Research questions

1. What are the qualitative differences in using hedges and boosters and their semantic and pragmatic subcategories between the Albanian and Italian L1 and L2 corpora?

2. What choices do students make in the usage of hedges and boosters: word order, orientation, manifestation, and prosody of modalization?

3. What are the quantitative differences in using hedges and boosters, and their semantic and pragmatic subcategories between the Albanian and Italian L1 and L2 corpora?

4. What are the quantitative differences in the usage of hedges and boosters: word order, orientation, manifestation, and prosody of modalization?
Research question 1:

What are the qualitative differences in using hedges and boosters and their semantic and pragmatic subcategories between the Albanian and Italian L1 and L2 corpora?

1.1 New introduced categories:

plausibility shields (e.g. *I think*)

attribution shields (e.g. *Sumra claims*) (Prince et al. 1980; Kaltenböck, Mihatsch & Schneider 2010)

impersonal shields (e.g. *it is suggested*),

multiple shields (e.g. *one of them is probably Ludovico di Breme, as Ungaretti himself suggests* [...] ((transl.V.DH from CIAO09MPL_7))

Clear inclusion / exclusion criteria

**Frequently excluded instances** for the lemma *prove*: *the desire of proving, failure to prove, has the burden to prove, wish /want /aim to prove, we should prove, let us prove, we have to prove*.

**Frequently included instances**: *we are now in position to prove, it suffices to prove and we start by proving*.

*my theoretical additions*
Research question 1:

What are the qualitative differences in using hedges and boosters and their semantic and pragmatic subcategories between the Albanian and Italian L1 and L2 corpora?

1.2 Overlap: The dual functions of different lexemes as hedges (suggest, believe, not entirely, not clearly) and boosters (strongly suggest, strongly/really/truly believe)

Subcategories: an overlap between the propositional and authorial categories within the same hedge and booster. Reflected in the quantitative analysis.

Similar to Varttala (cf. 2001: 11-12) and Skelton (cf. 1988: 38) and different from Prince et al. (cf. 1980: 6-20).

S (1) The internet is definitely the fastest-growing communication tool [...] (CARE13FBL_24).

S (2) [...] in order also to differentiate definitely one’s class from that class of nobility not able to follow the same rate of consumption [...] (CIAOE12FPS_13).

stronger colour=clearer booster
weaker colour=less booster
Research question 2:

What choices do students make in the usage of hedges and boosters: word order, orientation, manifestation, and prosody of modalization?

2.1 The interaction and overlap between modulation and modalization (ideational and interpersonal) within the same lexeme functioning in an area of uncertainty and negotiation. Modulation: Hedges.

**Modulation (+ideational)** proposing/recommending an idea, possible plan, or action (COD, QEP, FGJSSH)

S (4) *It is suggested to the teachers that these habits should be avoided in the classroom and proper use of nonverbal communication be used in order to have great impact on the teaching learning process.* (CARE15MML_7)

**Modalization (+interpersonal)** communicating an idea or feeling without stating it directly or giving proof (COD)

S (5) *This would perhaps suggest that, for example, that ‘Derived VC’ is actually an extreme form [...].* (Woodhall, 2003: 15; cited in Vázquez and Giner 2009: 229)
Research question 2:

What choices do students make in the usage of hedges and boosters: word order, orientation, manifestation, and prosody of modalization?

2.2 The non-applicability or the partial applicability of the phenomenon of 

transferred negation

(e.g. not clearly=vaguely) in Albanian and Italian (in contrast to English)

S (6) Prima della loro nascita esistevano alcune iniziative ma non chiaramente definibili WISE. (CIAO08FPE_4)

Before they were created, some initiatives existed, but they were not as clearly [vaguely] definable as WISE. (transl.V.DH)

S (7) La questione, che non riguarda certamente solo la metodologia ma che interessa trasversalmente ogni disciplina umana perché [...]. (CIAO10MPS_54)

The matter that certainly does not only have to do with the methodology but it also transversely interests every human discipline because [...]. (transl.V.DH)
Research question 3:
What are the quantitative differences in using hedges and boosters, and their semantic and pragmatic subcategories between the Albanian and Italian L1 and L2 corpora?

<table>
<thead>
<tr>
<th>corpus</th>
<th>logodds</th>
<th>n</th>
<th>proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAR</td>
<td>0.412</td>
<td>2445</td>
<td>0.536</td>
</tr>
<tr>
<td>CIAOE</td>
<td>-0.158</td>
<td>2737</td>
<td>0.375</td>
</tr>
<tr>
<td>CIAO</td>
<td>-0.254</td>
<td>2906</td>
<td>0.374</td>
</tr>
</tbody>
</table>

Table 3: The variable of corpus influencing the usage of hedges (and boosters).

Logistic Regression: H (hedge) or B (booster), run for H, remaining is B. Logodds: log \[p/(1-p)\], logarithm of the odds ratio of H, infinite +/- values, easily updated with new data.
Proportion: percentage of H, N values: total values of H and B (see also Gelman and Hill 2007; Chatterjee and Ali 2006)

Exclusion of CARE and other categories, empty cells, interaction

Italian appear more confident / fully committed in their L1 and L2 writings: boost
Albanians in L1 appear more tentative: hedge.

P: probab. of occurrence
P-1: probab. of non-occurrence
Research question 3:
What are the quantitative differences in using hedges and boosters, and their semantic and pragmatic subcategories between the Albanian and Italian L1 and L2 corpora?

3.1 Proud and confident Albanians, as generally known and Albanian university teachers also predicted lower vertical power role, institutional role and personal features of students in an academic context including more ‘powerful’ and experienced supervisors, teachers and other experts. (see also Argyle and Henderson 1984; Matthiessen 2010; Pulcini & Furiassi 2004) = more hedges by Albanians

Italians: confident, sometimes even over-confident, as Italian teachers predicted.
Research question 4:
What are the quantitative differences in the usage of hedges and boosters: word order, orientation, manifestation, and prosody of modalization?

4.2 Orientation - manifestation: hedges - significantly favored in the objective-implicit combination; boosters - frequently employed in the objective-explicit one.

<table>
<thead>
<tr>
<th>orient-manifest</th>
<th>logodds</th>
<th>n</th>
<th>proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>object.-implicit</td>
<td>0.138</td>
<td>6320</td>
<td>0.437</td>
</tr>
<tr>
<td>object.-explicit</td>
<td>-0.138</td>
<td>1768</td>
<td>0.376</td>
</tr>
</tbody>
</table>

Table 4: The single variable of orientation - manifestation influencing the usage of hedges (and boosters).

Objective-implicitness - tentative nature of hedging itself - the low institutional and power roles of students as actors in the academic discourse.

Objective-explicitness - boosters confirms the claims of the interviewed university teachers from Albania and Italy.
Research questions 3, 4

3. What are the quantitative differences in using hedges and boosters, and their semantic and pragmatic subcategories (a. commitment) between the Albanian and Italian L1 and L2 corpora?

4. What are the quantitative differences in the usage of hedges and boosters: word order, orientation, manifestation, and prosody of modalization?
References


References

References


A Corpus-based Comparison of Albanian and Italian Student Academic Writings in L1 and English as L2: Introducing Hedges and Boosters within Modalization, Modulation and Degree

Vincenzo Dheskali
Sixth Semester PhD Student
Chemnitz University of Technology
vincenzo.dheskali@s2015.tu-chemnitz.de
To appropriately convey plausible reasoning, refer to other sources, refer to our results, and use various realizations of modality and its metaphorical extension by balancing the right amount of conviction with tentativeness, it is at the heart of effective writing within the academic context. (cf. Halliday and Matthiessen 2014; Hyland 1998a; 1998b)

Those aspects should be considered by researchers focusing on hedges and the less researched boosters (Bondi 2008; Akbas and Hardman 2018) within the underexplored Albanian context (Toska 2015) and Italian context as well as university teachers in their classrooms.
Main variables and their categories

<table>
<thead>
<tr>
<th>Main variables</th>
<th>their categories/levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>word order</td>
<td>fronted, thematic, neutral, medial, final (afterthought), parenthetic, in parentheses</td>
</tr>
<tr>
<td>a. commitment</td>
<td>proposition-related (approximator for H, intensifier for B)</td>
</tr>
<tr>
<td>corpus</td>
<td>or author-related (shield for H, emphasizer for B)</td>
</tr>
<tr>
<td>orient-manifest</td>
<td>subjective-explicit, subjective-implicit, objective-explicit, objective-implicit</td>
</tr>
</tbody>
</table>

Table 5: The main variables related to hedges and boosters and their different levels or categories.

<table>
<thead>
<tr>
<th>current predictors</th>
<th>p. value</th>
</tr>
</thead>
<tbody>
<tr>
<td>word order</td>
<td>6.21e-64</td>
</tr>
<tr>
<td>a. commitment</td>
<td>8.92e-36</td>
</tr>
<tr>
<td>corpus</td>
<td>4.25e-33</td>
</tr>
<tr>
<td>orient-manifest</td>
<td>1.80e-04</td>
</tr>
</tbody>
</table>

Table 6: A general comparison of the significance of all variables influencing the usage of hedges (and boosters).