Transitivity Development in Spanish Learners of English

Mick O’Donnell
Universidad Autónoma de Madrid

Funded by Ministerio de Ciencia e Innovación (FFI2009-14436/FILO)
1. Background

- Work is part of the TREACLE Project.
- We use a corpus of written learner English from Spanish University students.
- Each essay is associated with the proficiency score of the learner.
- We study the corpus to better understand how a learner’s use of English develops as they progress in proficiency.
1. Background

Our interests at present are with grammatical development:

- **Errors** (grammar and vocab): (Murcia & MacDonald 2011)
- **Modality** (Garcia 2011)
- **Tense/Aspect** (O’Donnell 2012)
- **Transitivity** (this talk)
2. The TREACLE Project

- **Project**: TREACLE
- **Teaching Resource**
- **Extraction from an Annotated Corpus of Learner English**
- **A cooperation between:**
  - Universidad Autónoma de Madrid and Universitat Politècnica de Valencia
- **Runs**: January 2010 – December 2012

**Official Title**: “Developing an annotated corpus of learner English for pedagogical application”

**Funded by**: Ministerio de Ciencia e Innovación (FFI2009-14436/FILO)
3. The Corpus and Software

For this study, I used just the WriCLE corpus.

- **Size:** 709 essays of ~1000 words each (about 700,000 words)
- **Composition:** Written essays by Spanish learners of English at University level (Rollinson and Mendikoetxea 2008)
- **NOT all of the corpus used for this report:**

<table>
<thead>
<tr>
<th></th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
<th>Tot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essays</td>
<td>37</td>
<td>134</td>
<td>76</td>
<td>90</td>
<td>20</td>
<td>357</td>
</tr>
<tr>
<td>Words</td>
<td>24,000</td>
<td>94,000</td>
<td>52,000</td>
<td>113,000</td>
<td>21,000</td>
<td>304,000</td>
</tr>
<tr>
<td>Clauses</td>
<td>3,500</td>
<td>13,100</td>
<td>7,500</td>
<td>15,700</td>
<td>3,000</td>
<td>42,800</td>
</tr>
</tbody>
</table>
3. The **Corpus and Software**

- **UAM CorpusTool** used to automatically parse each text in terms of Transitivity analysis.

- Free: Macosx and Windows:
  - [http://www.wagsoft.com/CorpusTool/](http://www.wagsoft.com/CorpusTool/)

- Transitivity analysis only available in version 3.0, to be released in August

- Still buggy...
4. Methodology

1. Initial SVO analysis

- Clauses parsed by Stanford Parser (Klein and Manning 2003) to produce basic Subj^Verb^Obj analysis.

“I am happy to see you”

```
nsbj(happy-4, I-1)
cop(happy-4, am-2)
advmod(happy-4, always-3)
root(ROOT-0, happy-4)
aux(see-6, to-5)
xcomp(happy-4, see-6)
dobj(see-6, you-7)
```
4. Methodology

Can learner English be parsed reliably?:

- Actually, yes, with something like 80% reliability on each clause feature (some more, some less)
- This is enough to see trends.

Each level has its own problems:

- Low level learners make more lexical and grammar mistakes, which may throw the parser
- Higher level learners write better text but write longer sentences, which are harder for the parser to parse.
4. Methodology

2. Mood Analysis:

Stanford analyses mapped automatically to something closer to a Quirk and Greenbaum-style analysis:
4. **Methodology**

3. **Automatic Transitivity Analysis:**

   - The Mood analysis is used to derive a transitivity analysis of each clause unit:
   - Process type derived by:
     - a. Looking up verb in process-type lexicon (9,300 verb senses)
     - b. Where ambiguous, syntactic information used to disambiguate

![Table Example](image)
4. Methodology

3. Automatic Transitivity Analysis (ii):

Simple mapping rules used to map Mood roles onto Participant roles:

- If monotransitive verb in active clause:
  - Subj $\rightarrow$ Actor
  - DObj $\rightarrow$ Goal

- If monotransitive verb in passive clause:
  - Subj $\rightarrow$ Goal
  - by-DObj $\rightarrow$ Actor

<table>
<thead>
<tr>
<th>The Picaso</th>
<th>was bought</th>
<th>by them</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subj</td>
<td>Pred</td>
<td>DObj</td>
</tr>
<tr>
<td>Actor</td>
<td>Process</td>
<td>Goal</td>
</tr>
</tbody>
</table>

They bought the Picaso

Subj Pred DObj

Actor Process Goal
4. Methodology

Passive ditransitives more problematic:

- Mary was given a bouquet
- A bouquet was given to Mary

If DObj starts with ‘to’, Subj = Goal
Else: Subj = Recipient

Beneficial passives: Subj=Goal
(rare: Mary was built a house)
4. Methodology

Statistical Processing:

- Derived counts of use of each process type at each level of proficiency (5 levels in corpus)

- Derived counts of Participant role sequences for each process type, e.g.,
  - verbal-passive-addressee-process-sayer-phenomea
    *(She was told by me to go)*
5. Results (i): General Process Type Usage

Changing mix of process type usage with increasing proficiency: doesn’t seem like much, but some shifts: fall in relational, increase in verbal
5. Results (i): General Process Type Usage

- **Material**: A2 > B1 > B2 > C1 > C2
- **Verbal**: A2 = B1 > B2 > C1 > C2
- **Mental**: A2 > B1 > B2 > C1 > C2
- **Relational**: A2 > B1 = B2 > C1 > C2
5. Results (ii): Material Processes

- Three classes of material verbs:
  - Ergative (state change, some motion)
  - Monotransitive (acting on)
  - Ditransitive (exchanging, acting for)
- Main change: ergative verbs used less with increasing proficiency.

![Bar chart showing ergative usage](chart.png)

![Pie chart showing verb usage](chart.png)
5. Results (ii): Material Processes

Change in usage of different ergative patterns:

- General fall in the "I sailed the boat" type construction,
- An increase in the "The boat was sailed" type construction.
- Probably reflects the more general move towards passive forms.
- No real change in use of the middle form (The boat sailed)
5. Results (ii): Material Processes

- Ditransitive verbs in active clauses:
  - Most ditrans. verbs used with just two participants.
  - Little realisation of Recipient
  - No realisation of Beneficiary
5. Results (ii): Material Processes

- Ditransitive verbs in Passive clauses:
- As with other process types, increased use of passive with ditransitive verbs
- Most of increase in Recipient^Process^Goal structures (Mary was given an apple)
5. Results (iii): Verbal Processes

Verbal Processes:

- General increase in verbal processes in comparison with other processes - learning to quote!
- No clear change in addressee-oriented vs. non-addressee oriented:
  
  (I told John that... vs. I said that...)

- Other evidence (from Error Analysis) that Spanish learners often use “say” type verbs with an Addressee, e.g., “He say me to go”.

5. Results (iii): Verbal Processes

Verbal Passives: very clear increase in passive with verbal processes! Up to 26%!!!

Main increase in “It could be argued that...” type structures (postponed Verbiage Subject)

Students learning to distance themselves from their claims.
5. Results (iv): Mental Processes

Mental processes:

- As with other processes, clear increase in passive forms:
  - *It is considered/believed/expected/felt that* ...
    (postponed Subj=Phenom.)

- Again, students avoiding mention of the Senser!
5. **Results (v): Relational Processes**

- **Relational processes:**
  - No results currently.
  - System does not currently recognise subtypes of Relational Processes.
  - For future work.

- But general fall in use of relationals in respect to other process types.

- No idea why!
5. Results (vi): Existential Processes

- **Existential processes**: Includes structures like:
  - “There is a problem.”
  - “On the desk were some folios”

No major change in use with development

This structure can be transferred fairly directly from Spanish (*Hay un problema*) even by low level learners.
6. **Summary**

- This talk has presented a methodology for exploring how learners develop their transitivity resources as they progress in proficiency.
- Results may suffer due to accuracy of the parser (exact accuracy still needs to be established!)
- However, some clear patterns appear, particularly:
  - Increased use of verbal processes shows students are learning to introduce other voices.
  - Increased use of passive shows hiding of agency increasing.
  - This is particularly true for verbal and mental processes, and use of postponed subjects.
6. **Summary (ii)**

- The other goal of this paper was to introduce the first SFL Transitivity parser which is (will soon be) freely available.
- Built into next release of UAM CorpusTool.
- Accuracy will improve over time.
7. Future Work

- The accuracy of the automatic analysis needs to be improved:
  - Stanford parser is continuing to be tweaked by the Stanford team to improve its analysis.
  - Mapping from the Stanford parse to my Mood level needs to better capture some rare structures.
  - The process type lexicon is slowly being edited to contain more verbs and to remove mistaken or rare senses.
  - Mapping from Mood to Transitivity still needs some work.