

Transitivity Development in Spanish Learners of English

*Mick O'Donnell
Universidad Autónoma de Madrid*

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



Official Title: "Developing an annotated corpus of learner English for pedagogical application"

- 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
- 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:

	<i>A2</i>	<i>B1</i>	<i>B2</i>	<i>C1</i>	<i>C2</i>	Tot.
<i>Essays</i>	37	134	76	90	20	357
<i>Words</i>	24,000	94,000	52,000	113,000	21,000	304,000
<i>Clauses</i>	3,500	13,100	7,500	15,700	3,000	42,800

3. The Corpus and Software

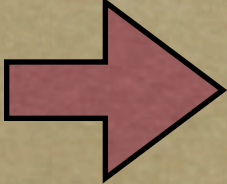
- UAM CorpusTool used to automatically parse each text in terms of Transitivity analysis.
- Free: Macosx and Windows:
 - <http://www.wagsoft.com/CorpusTool/>
- Transitivity analysis only available in version 3.0, to be released in August
- Still buggy...

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”



```
nsubj(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:

Grammar analysis for: Texts/Newt-Mittens.txt

Gingrich launched a blistering attack on Romney in what is essentially the h

Subj	Pred	DObj			Adjunct					
Thing		Det	PreMod	Thing	Op	Complem	Op			
			Head			Thing		Subj	Pred	Adjunct
										Det

<< < > >> Ignore Delete Other Action... Save Close Help

Assigned	TENSE-ASPECT	Gloss
not-progressive-aspect	simple-present	
grammatical-unit	simple-past	
simple-finite-clause	simple-future	
declarative-clause	simple-modal	
material-clause	present-perfect	
transitive clause	past perfect	

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

<i>Gingrich launched a blistering attack on Romney in what is essentially</i>					
Actor	Process	Goal	Circumstance		
				Carrier	Process
					Circumstance

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 → Actor

- DObj → Goal

<i>They</i>	<i>bought</i>	<i>the Picasso</i>
<i>Subj</i>	<i>Pred</i>	<i>DObj</i>
<i>Actor</i>	<i>Process</i>	<i>Goal</i>

- If monotransitive verb in passive clause:

- Subj → Goal

- by-DObj → Actor

<i>The Picasso</i>	<i>was bought</i>	<i>by them</i>
<i>Subj</i>	<i>Pred</i>	<i>DObj</i>
<i>Actor</i>	<i>Process</i>	<i>Goal</i>

4. Methodology

- Passive ditransitives more problematic:

- Mary was given a bouquet*

<i>Mary</i>	<i>was given</i>	<i>a bouquet</i>
<i>Subj</i>	<i>Pred</i>	<i>DObj</i>
<i>Recipient</i>	<i>Process</i>	<i>Goal</i>

- A bouquet was given to Mary*

<i>A bouquet</i>	<i>was given</i>	<i>to Mary</i>
<i>Subj</i>	<i>Pred</i>	<i>DObj</i>
<i>Goal</i>	<i>Process</i>	<i>Recipient</i>

- If DObj starts with 'to', Subj = Goal
- Else: Subj = Recipient
- Beneficial passives: Subj=Goal
- (rare: *Mary was built a house*)



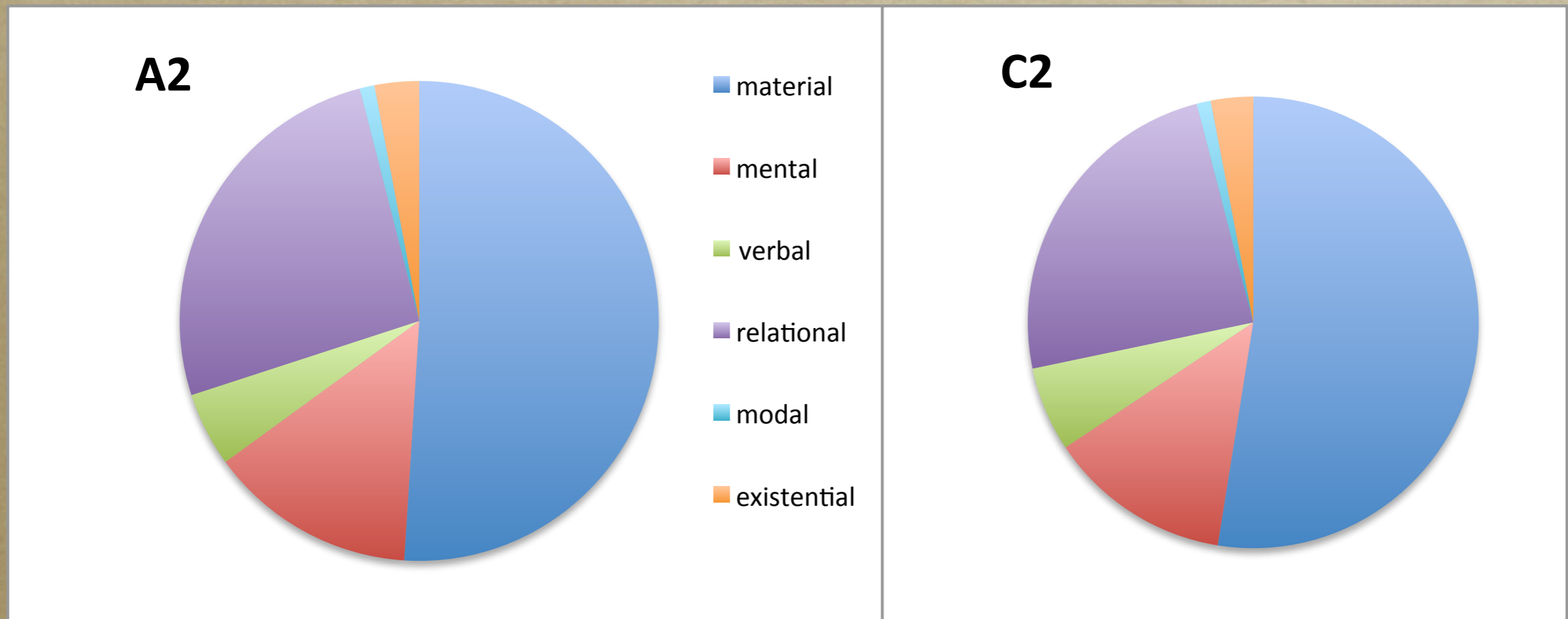
4. Methodology

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)

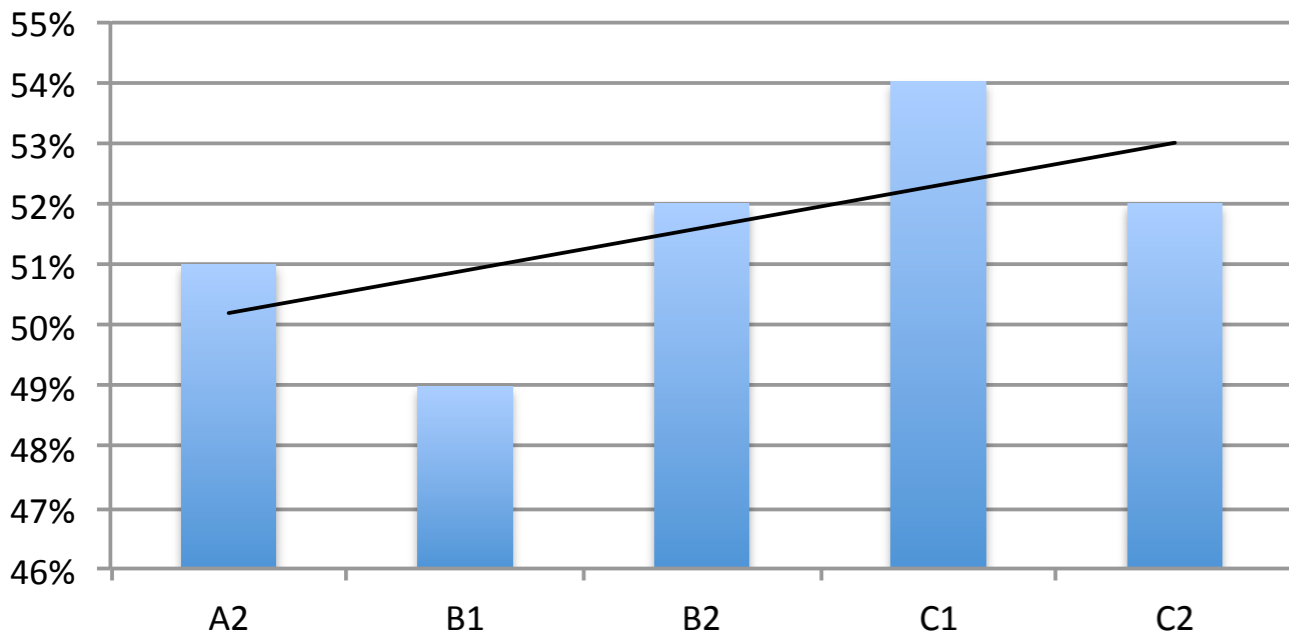
5. Results (i): General Process Type Usage

- Changing mix of process type usage with increasing proficiency: doesn't seem like much, but some shifts: fall in relational, increase in verbal

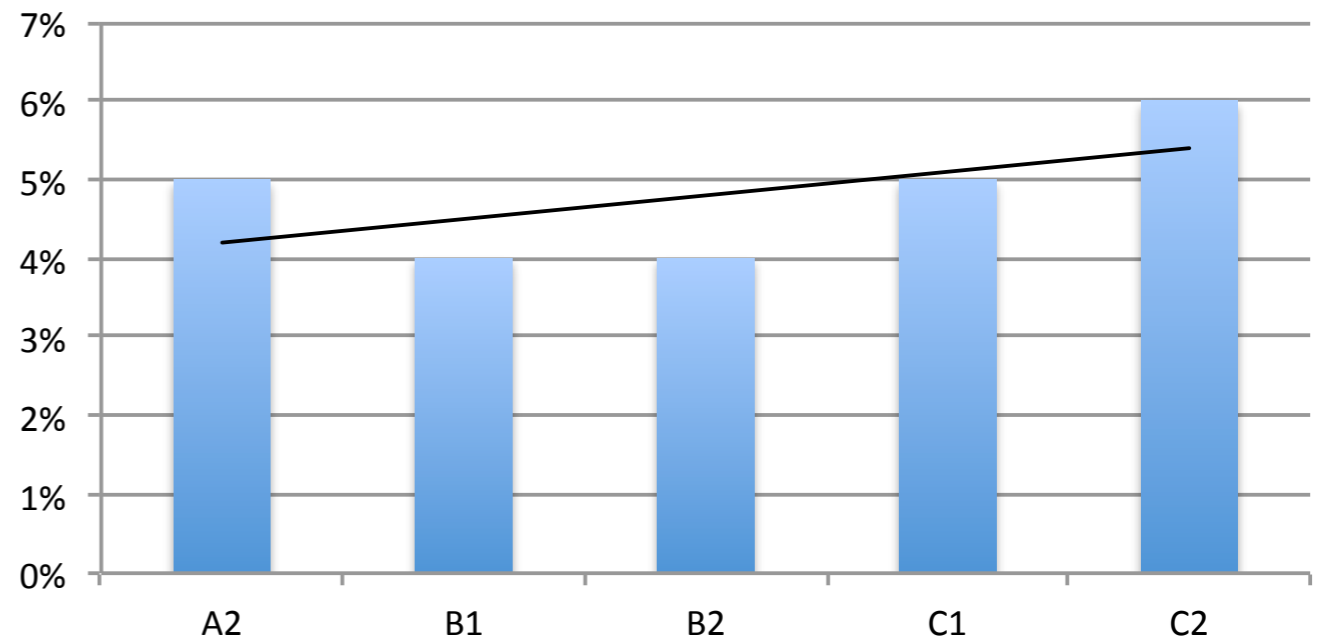


5. Results (i): General Process Type Usage

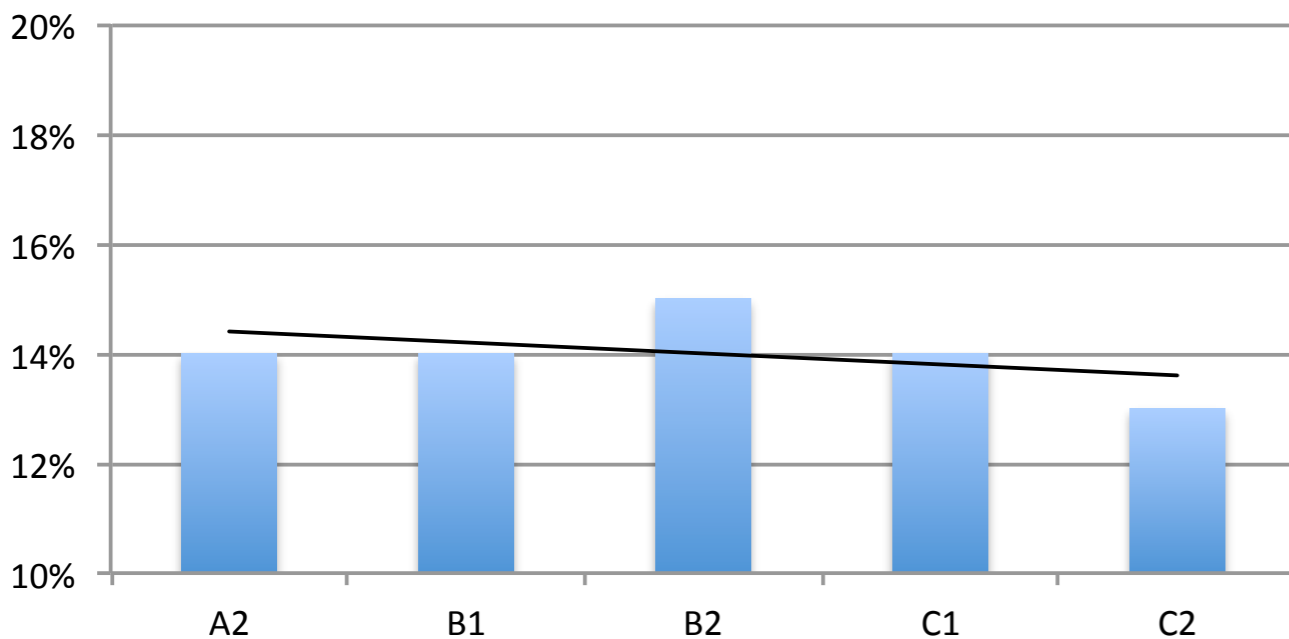
material



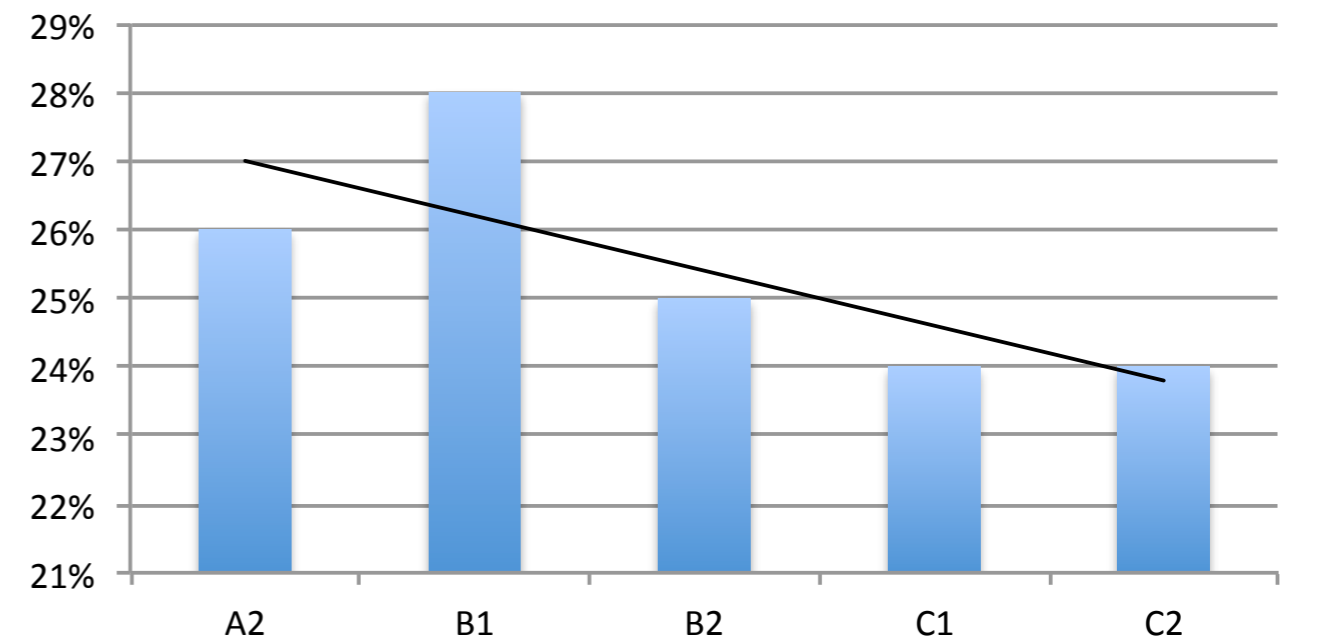
verbal



mental

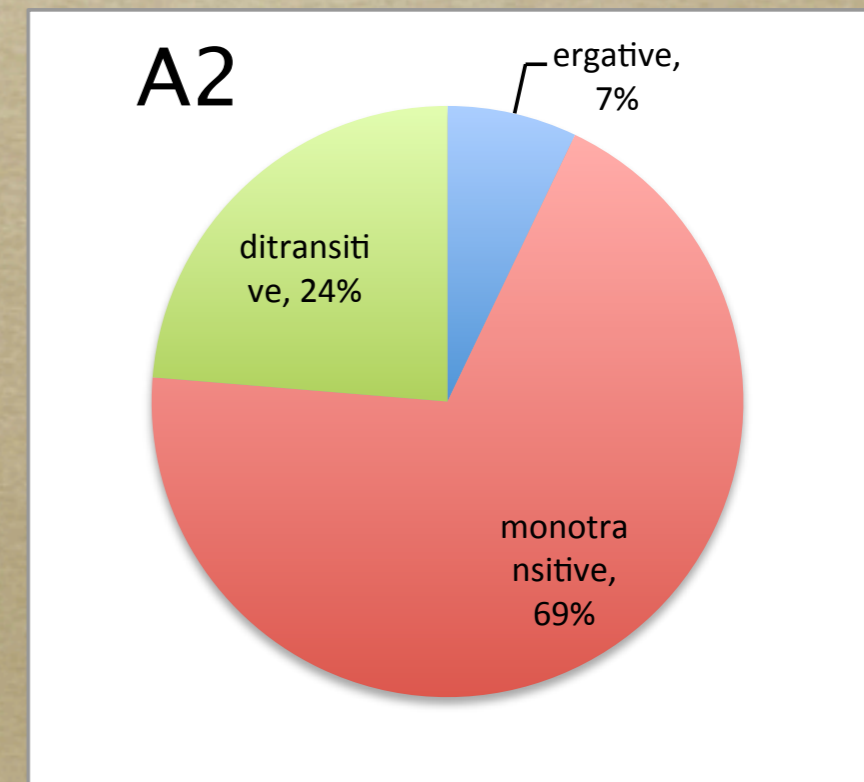
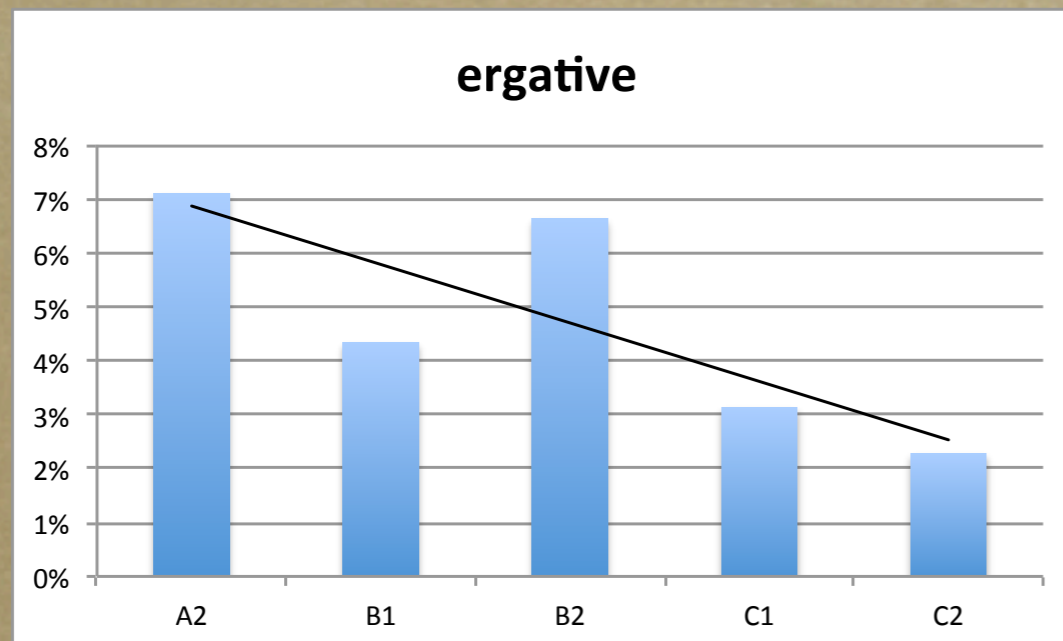


relational



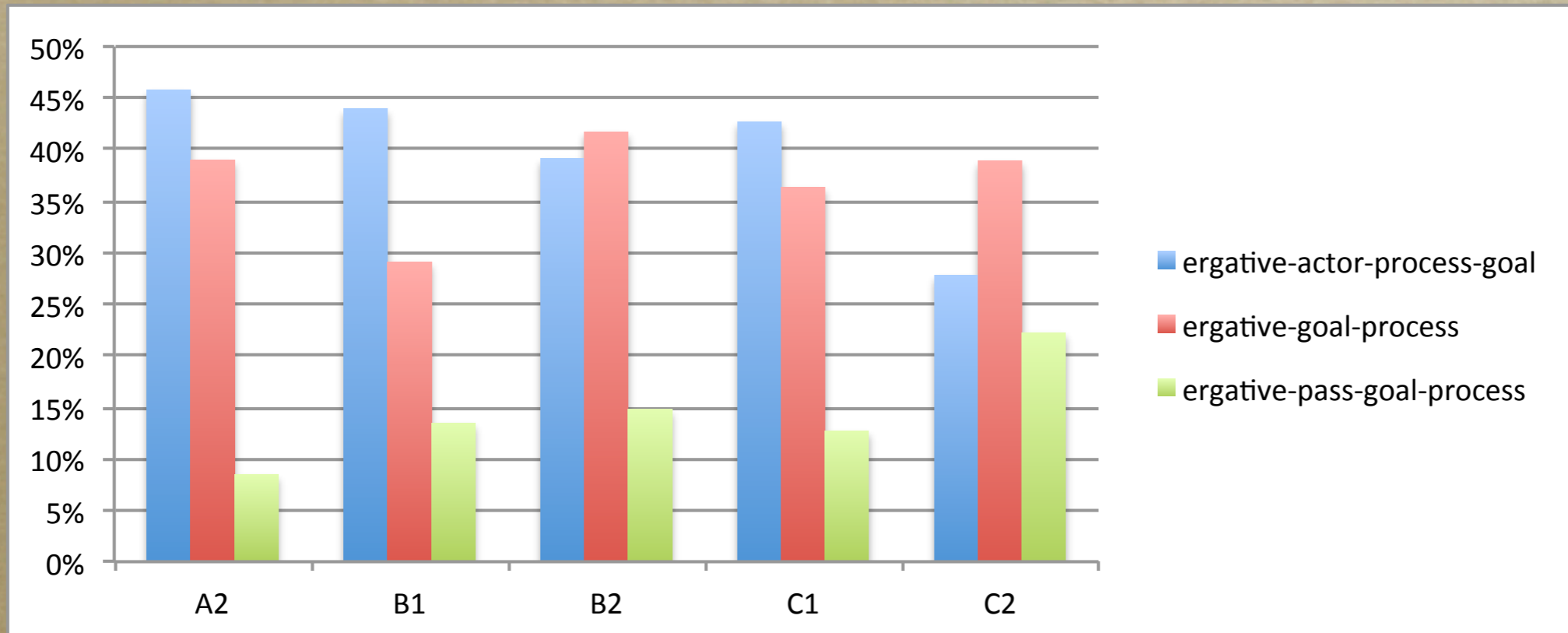
5. Results (ii): Material Processes

- 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.



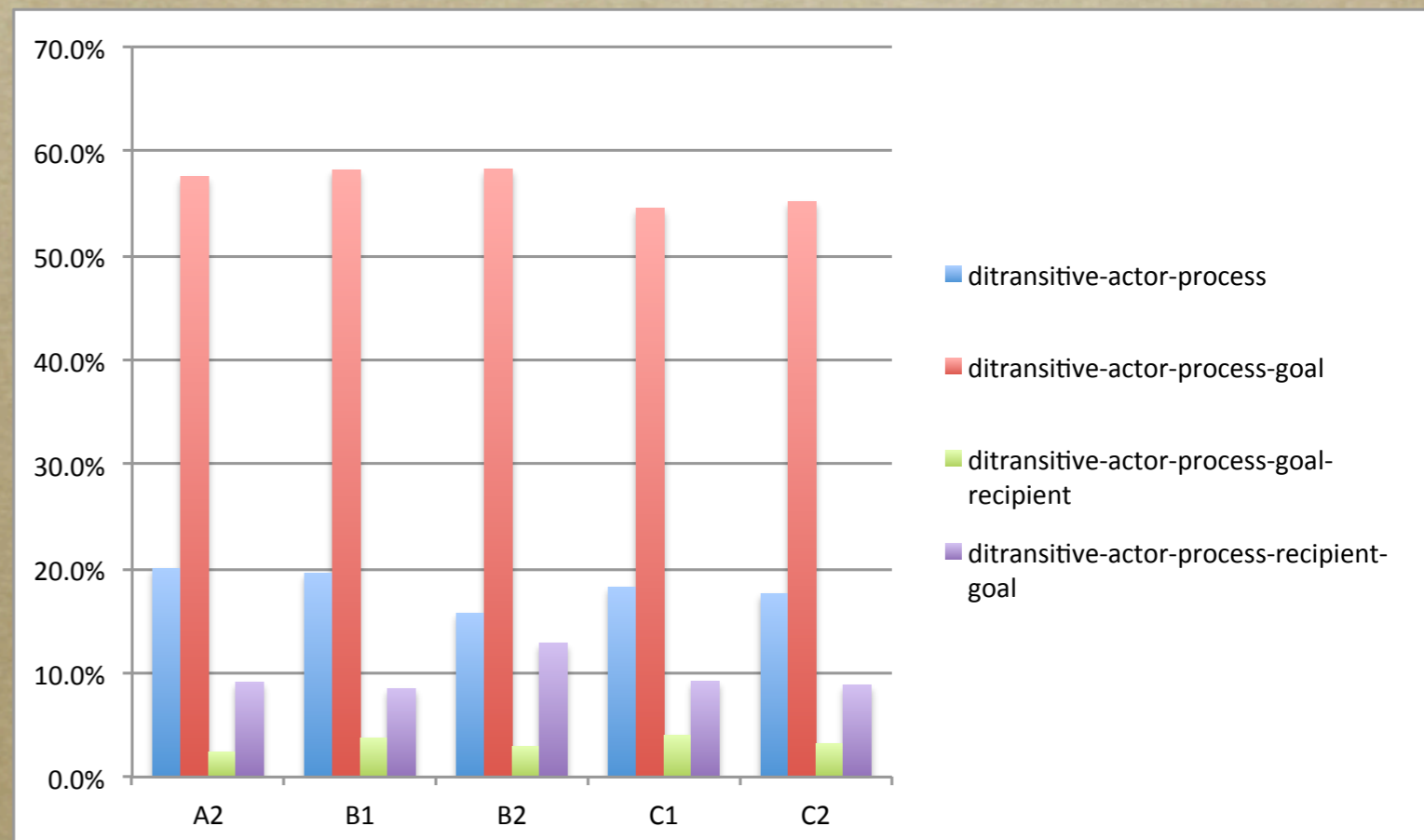
5. Results (ii): Material Processes

- 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*)



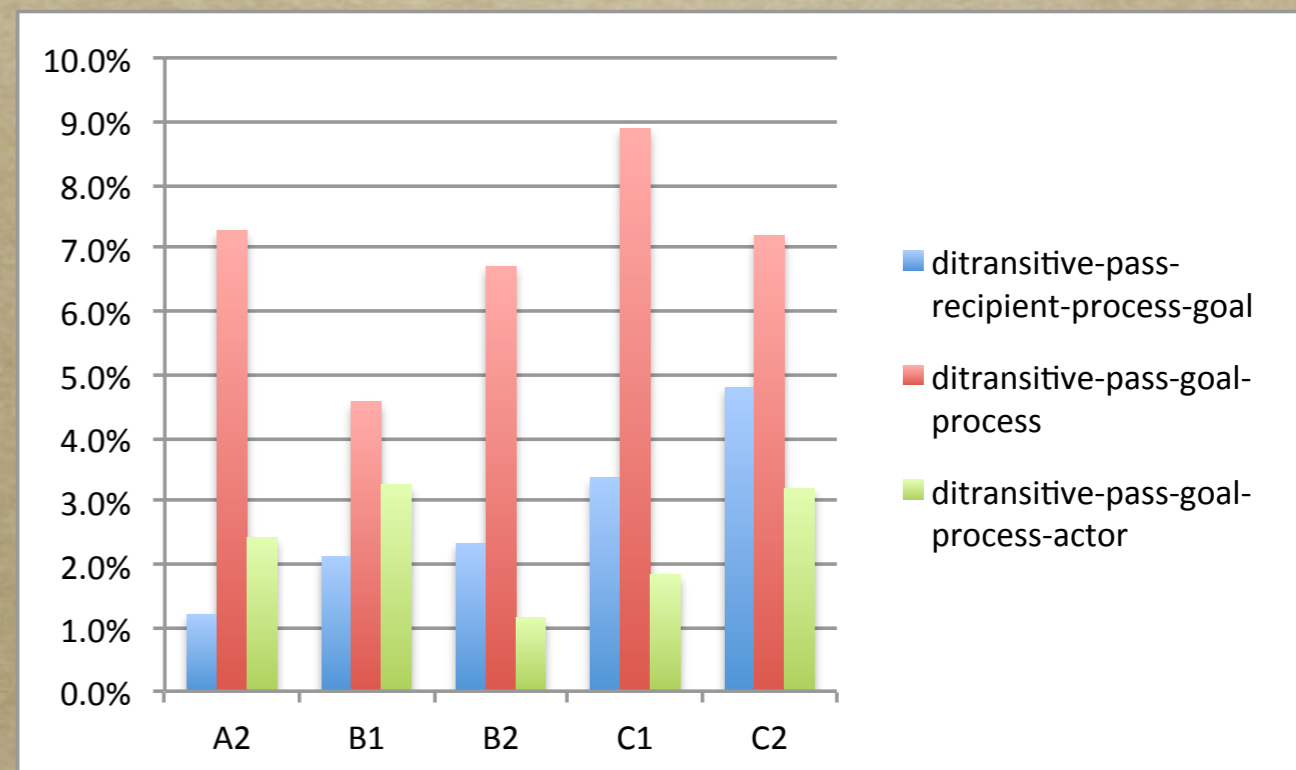
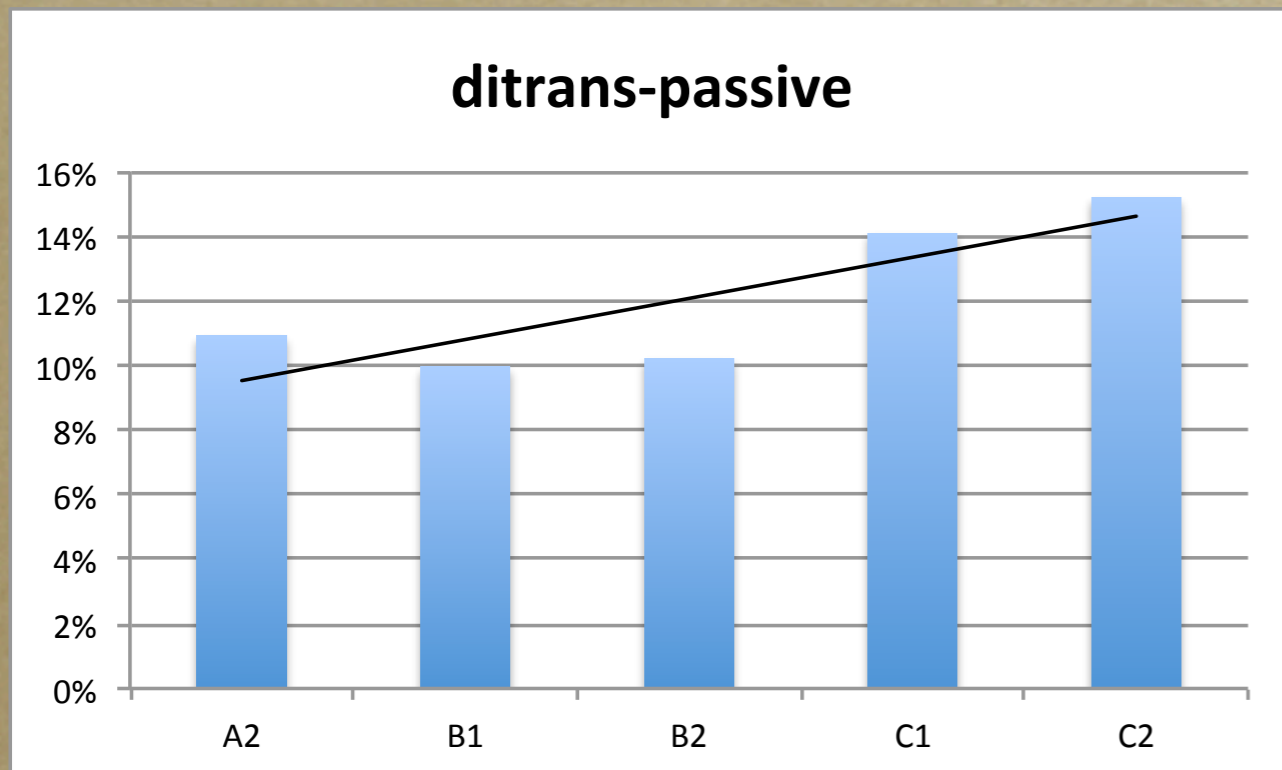
5. Results (ii): Material Processes

- 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*)



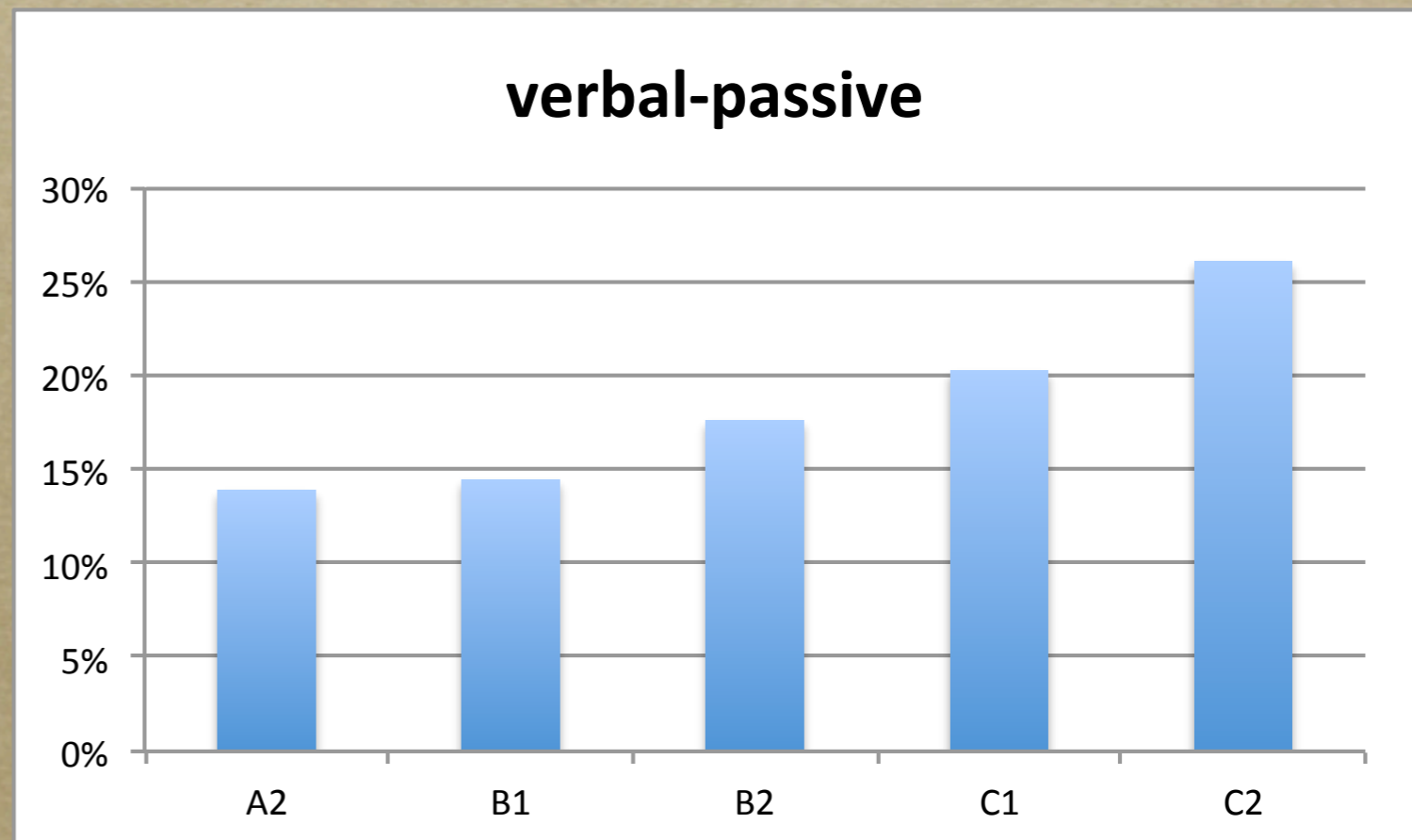
5. Results (iii): Verbal Processes

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.

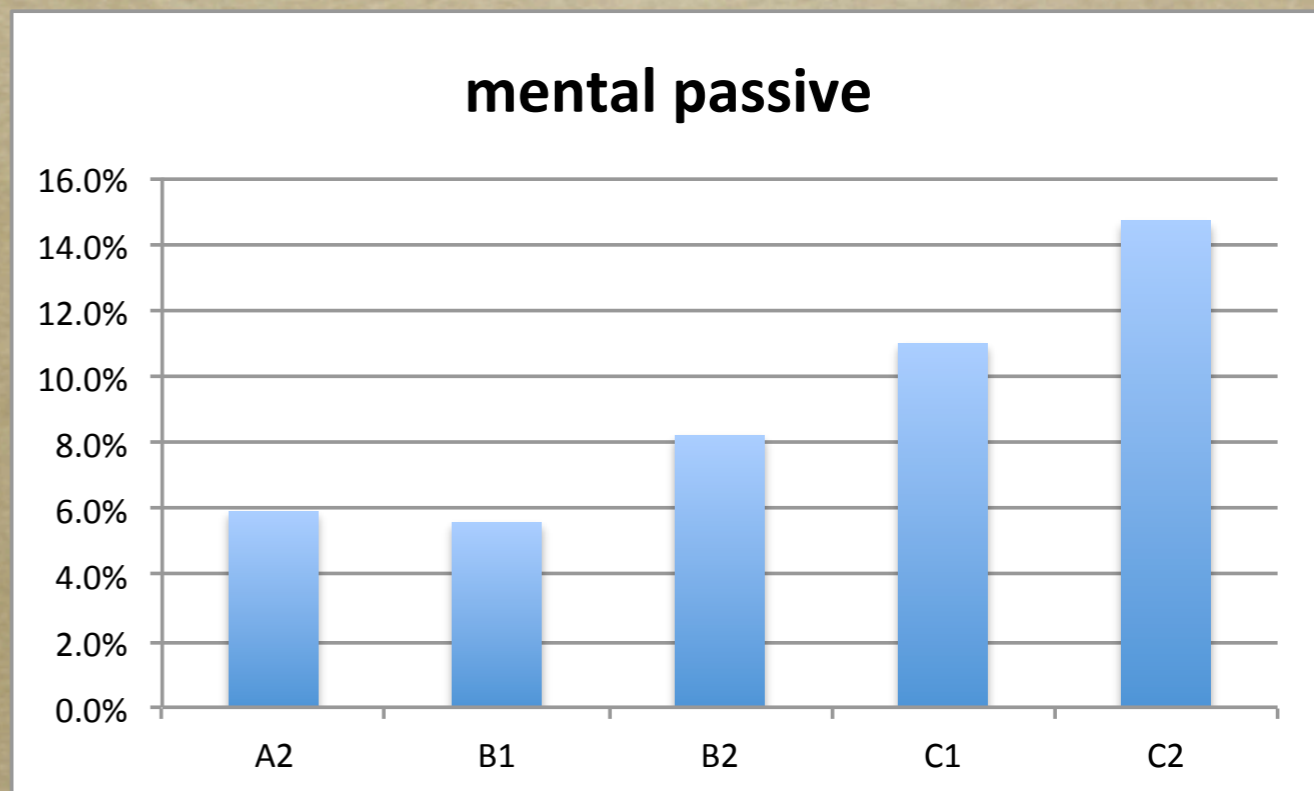


5. Results (iv): Mental Processes

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!

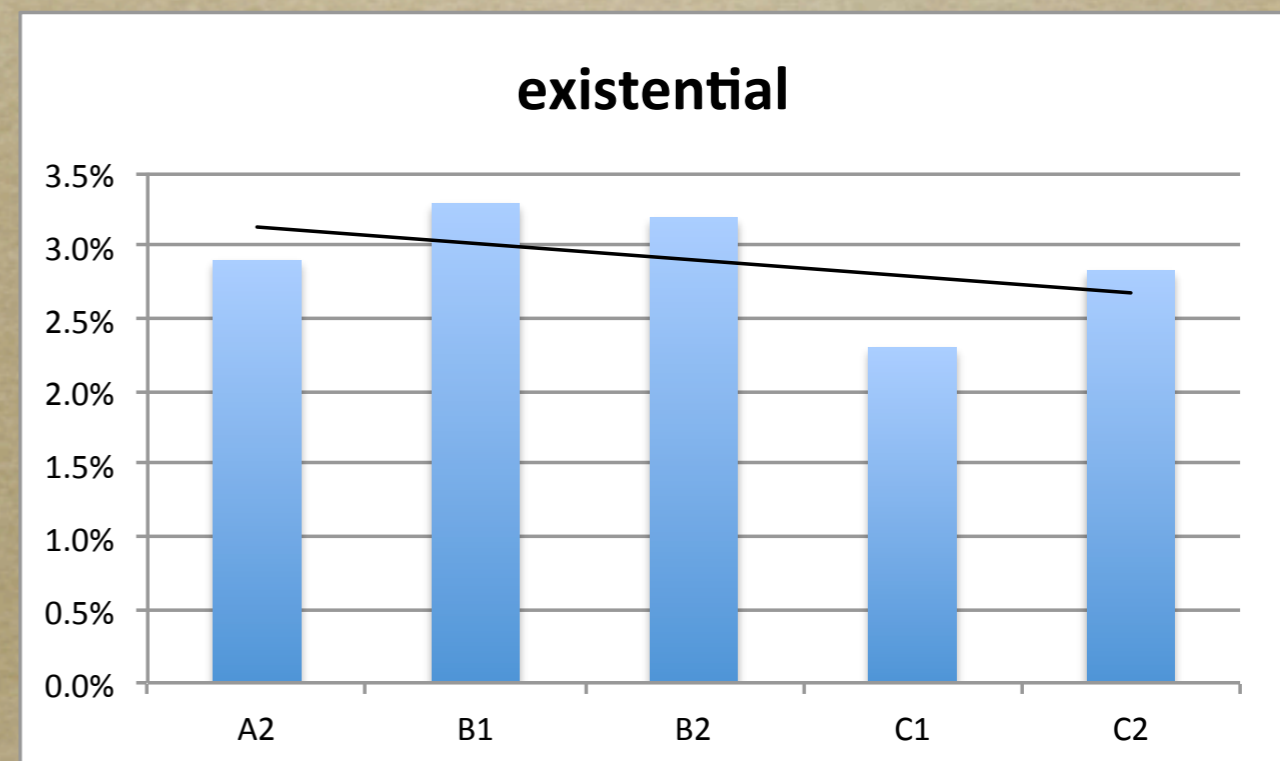


5. Results (v): Relational Processes

- 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