Classroom-based learner corpora: an empirical bridge between second language acquisition research and teaching practice

Dora Alexopoulou
EF Research Lab for Applied Language Learning
Linguistics Section, Faculty of Modern and Medieval Languages
Cambridge University

Language analysis to enhance language teaching
University of Leeds
July 2019

Learner Profiles: research questions

- Description: who does what when?
- Explanation: why?
- Intervention: how?

Second language acquisition in a real life context

Applied language learning

Big Educational Data

- Lots of data
- Diverse data: students, geographical context, teaching settings, teachers, etc.

Scaling up research from lab conditions to the real world

- Exploring complexity
- Developing richer models and deeper theories

The EF Cambridge Open Language Database EFCAMDAT

Size

- 177,000 learners (172 nationalities)
- 1.2 million writing samples
- 71 million words

Teacher input

36% of scripts have teacher corrections

Tasks

128 writing tasks

Proficiency

16 CEFR-aligned proficiency levels

Individual data

Scope for analysis of individual progress over time.
Nationality is crossed with country of access to EF to provide National Language as a proxy to L1 background.

Language and Communication

- Grammar is a code facilitating communication.
- The core features of this code are independent of communication and culture in the way the chemical properties of wine are independent of its social and cultural function.

Language and Communication

- Grammar is a code facilitating communication.
- The core features of this code are independent of communication and culture in the way the chemical properties of wine are independent of its social and cultural function.

From L1 to L2

- **Universals of language structure**: the basic principles of how words and phrases are put together in sentences
- **Variation between languages**: sounds, form marking (e.g. verb endings, gender, articles etc.)

(Foyle and Flynn 2013, Jarvis and Pavlenko 2007, White 1989)

What is free and what is challenging L2

- L2 learners are efficient communicators, often with mature cognitive systems (adults vs. kids).
- They can transfer universal aspects of linguistic structure (e.g. types of sentences).
- Language specific features: similarity between native language and English crucial: typologically similar features easy, typological differences result in challenges and negative transfer.
- As efficient communicators and with advanced cognitive capacities L2 learners produce early on complex language, but inaccurate—with errors.

(White,1989,Klein and Perdue 1989, Schepens et. al 2016, Slabakova 2009,2014,Scwartz and Sprouse 1996)

Relative Clauses

(1) This job is really the most suitable job what I have found for you.

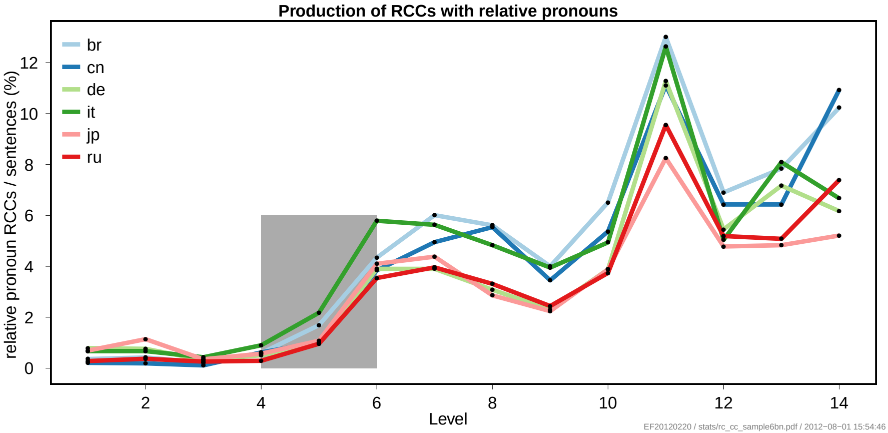
(2) I had to married an awful man thatI don't love for some time...

(3) If you want to know opinion that what you need.

(4) the e-ticket is a receipt what you paid your ticket .

(5) you shouldn't pay lots of money for things what you don't need

Learners use relative clauses before they are introduced in teaching



National language patterns

- Germans, Chinese and Russian learners overuse who-relatives and avoid that-relatives.

(6) a man **that/who** I don’t love any more

(7) a singer-song writer **that/who** brought your feelings on this music

Why early linguistic complexity?

(8) I had to married an awful man that I don’t love for some time...

(9) I had to marry an awful man (and) I don’t love the awful man for some time...

(10) I had to marry an awful man (and) I don’t love him for some time...

(11) I had to marry a man (and) the man is awful (and) I don’t love the man for some time....

7 more complex than 8 more complex than 9 complex than 10

10 = 18 words (without ‘and’) vs. 7 = 14 words

Why early linguistic complexity?

(8) I had to married an awful man that I don’t love for some time...

(9) I had to marry an awful man (and) I don’t love the awful man for some time...

(10) I had to marry an awful man (and) I don’t love him for some time...

(11) I had to marry a man (and) the man is awful (and) I don’t love the man for some time....

7 more complex than 8 more complex than 9 complex than 10
10 = 18 words (without ‘and’) vs. 7 = 14 words

Why early linguistic complexity?

(8) I had to married an awful man that I don’t love for some time...

(9) I had to marry an awful man (and) I don’t love the awful man for some time...

(10) I had to marry an awful man (and) I don’t love him for some time...

(11) I had to marry a man (and) the man is awful (and) I don’t love the man for some time....

7 more complex than 8 more complex than 9 complex than 10
10 = 18 words (without ‘and’) vs. 7 = 14 words

Why early linguistic complexity?

(8) I had to married an awful man that I don’t love for some time...

(9) I had to marry an awful man (and) I don’t love the awful man for some time...

(10) I had to marry an awful man (and) I don’t love him for some time...

(11) I had to marry a man (and) the man is awful (and) I don’t love the man for some time....

7 more complex than 8 more complex than 9 complex than 10
10 = 18 words (without ‘and’) vs. 7 = 14 words

Why early linguistic complexity?

(8) I had to married an awful man that I don’t love for some time...

(9) I had to marry an awful man (and) I don’t love the awful man for some time...

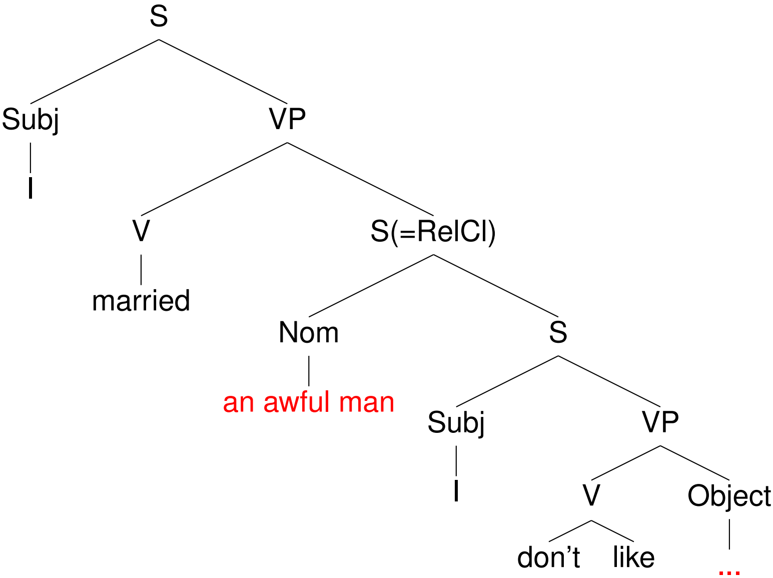
(10) I had to marry an awful man (and) I don’t love him for some time...

(11) I had to marry a man (and) the man is awful (and) I don’t love the man for some time....

7 more complex than 8 more complex than 9 complex than 10

10 = 18 words (without ‘and’) vs. 7 = 14 words

How can they do it?



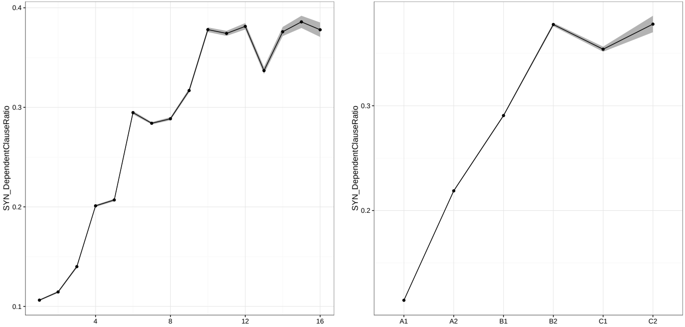


Figure: Subordinate Clause ratio across Englishtown levels (left) and CEFR-aligned (right)

Alexopoulou, T., Michel, M., Murakami, A. and Meurers, D. (2017).

Language specific features: similarity between mother-tongue and L2

- Learners can transfer features from their L1 to their L2 when the two languages are typologically similar.
- Learners from languages without articles show persistent difficulties with the article until very late proficiency.
- Verbal morphology is challenging for e.g. Chinese learners.
- These problems persist even in conditions of rich input (immersion) and early childhood start in language learning.

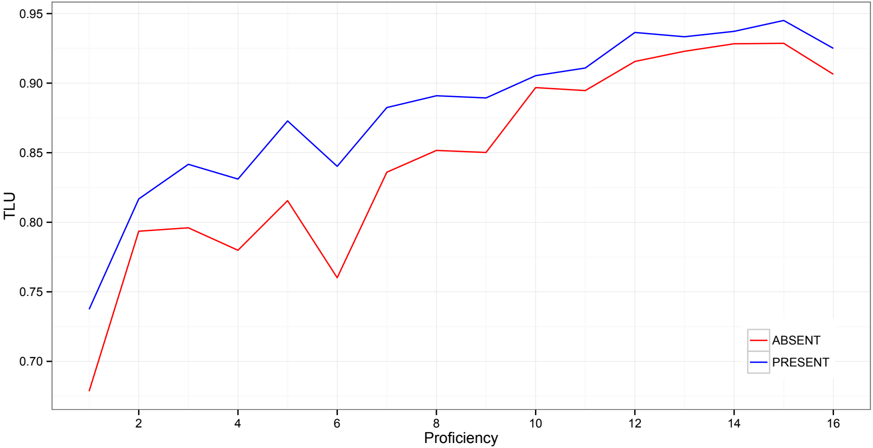
Scheppens et al. 2016, Paradis,J. 2017, Lardiere 1998

Language specific features: similarity between mother-tongue and L2

- Learners can transfer features from their L1 to their L2 when the two languages are typologically similar.
- Learners from languages without articles show persistent difficulties with the article until very late proficiency.
- Verbal morphology is challenging for e.g. Chinese learners.
- These problems persist even in conditions of rich input (immersion) and early childhood start in language learning.

Scheppens et al. 2016, Paradis,J. 2017, Lardiere 1998

Learner Profiles I
The English Article



Murakami, A. and Alexopoulou, T. (2016).

Where do errors come from?

As learners process the input they form generalisations or hypotheses about how forms might be linked to features that are not present in their mother tongue or are organised in a different way.

Errors are **not** random; they are creative and systematic.

Errors look messy because learners try different hypotheses and form-meaning mappings as their learning progresses and there is a lot of variation between learners as they get 'stuck' to different strategies and hypotheses.

Where do errors come from?

As learners process the input they form generalisations or hypotheses about how forms might be linked to features that are not present in their mother tongue or are organised in a different way.

Errors are **not** random; they are creative and systematic.

Errors look messy because learners try different hypotheses and form-meaning mappings as their learning progresses and there is a lot of variation between learners as they get 'stuck' to different strategies and hypotheses.

Where do errors come from?

As learners process the input they form generalisations or hypotheses about how forms might be linked to features that are not present in their mother tongue or are organised in a different way.

Errors are **not** random; they are creative and systematic.

Errors look messy because learners try different hypotheses and form-meaning mappings as their learning progresses and there is a lot of variation between learners as they get ’stuck’ to different strategies and hypotheses.

Uses of indefinte definite, indefinite and zero article

1. **def article**: *I like the book/the books*.
2. **indefinite article**: *I bought a book*.
3. **zero article**: *I like books with nice covers*. *I like sugar in my coffee*.

Uses of indefinte definite, indefinite and zero article

1. **def article**: *I like the book/the books*. **38%**
2. **indefinite article**: *I bought a book*. **31 %**
3. **zero article**: *I like books with nice covers*. *I like sugar in my coffee*. **31%**

a or zero?

- **Existential**: There is a problem. We have a problem. vs. There are problems. We have problems.
- **Referential**: Last night we watched a movie. vs. Last night we watched movies and listened to music.
- **Non-referential:** I need a toothbrush. You can get a nice coat online. vs. I don’t take sugar in my tea. You can find lovely flowers at Homebase. They sell nice notebooks at Waterstone’s.
- **Predicative**: She is a doctor. She is a nice person. vs. They are maths teachers. These are peaceful demonstrations.
- **Kind**: Lions are wild animals. Bees are threatened with extinction. They invited only women. Smart phones have become a necessity.
- **Idiomatic**: It was such a good party. vs. She goes to bed early.

Errors

Article omission most common error for all learners.
Learners doing well with idiomatic uses (over 80% accuracy).
Most challenging use/errors:
Non-referential indefinites: omission and overuse of 'the' instead of 'a'. Increased error with abstract nouns.

1. *It should be place where I can get great experience (B1, L1 Russian).*
2. *Well, e-ticket is a kind of receipt that shows you already bought your travel ticket. (A2, L1 Brazilian).*
3. *I will find good job.*
4. *I think every progress occurs after solving a problem and an experience comes after the moment when you really need it.*
5. *last week I visited a museum of the art in a small, old city. (B1, L1 Brazilian)*

(Derkach, K, phd thesis in preparation)

Causes of article errors

- Hypothesis: learners associate the indefinite article with specificity and concretness instead of number and the mass/count distinction.
- Increased ommission with non-referential and predicative uses and abstract nouns.

Brazilians: overuse of the definite article with generics

Certainly, here in Brazil we have some of the strictest laws of the world regarding the smoking.Obviously, it is recent. About ten years ago, or a little bit more, the advertising promoted the smokinghabits using lifestyle, freedom, adventure and, several times, health and sports. Today it is unthinkable. But at that moment, that was a most effectivestrategy to sell smoking cigarettes in my country. However, today is very different. The advertising is very restricted. The communication must be made to adult people. The smoking industry cant associate your products and brand to sports, health or lifestyle. The packing of cigarettes must show the effects of smoking habits, such as a cancer and other diseases. Furthermore, smoke is prohibited at several public places, restaurants and shopping. I agree with restrictions. I think today is better than before that laws.

Teaching interventions: Input Instruction

- Aim to change the way learners process input so as to help them establish the right form-meaning mappings (Van Patten 1996).
- Interventions drawing learners attention to number marking. (Derkach, PhD thesis in preparation).

Conclusions

- Language analysis for identifying patterns in learner production and explaining the underlying generalisations and hypotheses learners use during their learning.
- Learner profiles across proficiency and L1 backgrounds—crosslinguistic influence can inform who to target, when and how.
- Learner corpora from educational settings an empirical bridge between lab-based SLA research and teaching practice.