Beyond counters in the head: uncertainty and negative evidence in L2 lexical learning

Cylcia Bolibaugh

Centre for Research in Language Learning and Use

UNIVERSITY of York

Introduction

- A heteromorphic distributed lexicon:
  - *shed light, monumental error, low cost*
  - *the* X-*er, the* Y-*er; NP be-TENSE sorry to keep-TENSE NP waiting; it's more a question of NP than of NP*
  - *never mind, don't mention it, come to think of it*
  - *going great guns, as daft as a brush*

- A second language idiomatic deficit :

Traffic-aggression?
Law-obedient people?
Slim escape/narrow chance?

(e.g. Granger 1998, Howarth 1998, Wray 2002, Nesselhauf 2004, Siyanova & Schmitt 2007, 2008)

Regions
The Strength To Help You Do More.

Dear client of Regions Bank,

Technical services of the Regions Bank are carrying out a planned software upgrade.

We earnestly ask you to visit the following link to start the procedure of confirmation on customers data.

To get started, please click the link below:
https://online.regions.com/ibsregions/cmserver/users/default/confirm.cfm

This instruction has been sent to all bank customers and is obligatory to follow.

Thank you,

Customer Support Service.

This message contains graphics. If you do not see the graphics, click here to view.

Regions
The Strength To Help You Do More.

Dear client of Regions Bank,

Technical services of the Regions Bank are carrying out a planned software upgrade.

We earnestly ask you to visit the following link to start the procedure of confirmation on customers data.

To get started, please click the link below:
https://online.regions.com/ibsregions/cmserver/users/default/confirm.cfm

This instruction has been sent to all bank customers and is obligatory to follow.

Thank you,

Customer Support Service.

Fundamentally different processes in L1 and L2 (e.g. Durrant and Schmitt 2010, Wolter and Gyllstad 2015)?

"for some reason, learners do not pay attention to collocational relationships" (Wray 2002: 183).

"Objects once experienced together tend to become associated in the imagination, so that when anyone of them is thought of, the others are likely to be thought of also" (James 1890, cited in Ellis 2001:42).

Associative model of phraseological learning

(e.g. Ellis 1997, Hoey 2005)

UNIVERSITY of York

Fundamentally different processes in L1 and L2 (e.g. Durrant and Schmitt 2010, Wolter and Gyllstad 2015)?

+ L2 Word level analytical bias

+ L2 Lack of surface level, syntagmatic encoding

+ dual route model of processing: default holistic processing in L1, analytic in L2

A lack of incremental learning?

- Durrant & Schmitt (2010): timed explicit recall of collocations met once in the input.

- Some processing advantages for L2 FS (Siyanova-Chanturia et al 2011;Wolter & Gyllstad 2011, 2013)

A question of amount & type of input?

Implicit tallying & appropriate association

"adult learners who implicitly retain memories of the words that go together in their input ought to establish strengths of association which are appropriate to that input" (Durrant and Schmitt 2010: 169).

while this is true for monolinguals, it is not (necessarily) so for bilinguals

Incidental SLA marked by a disassociation between learning what *is* acceptable & what *is not*.

Increases in knowledge of appropriate word combination is not accompanied by a decrease in false positives (production and acceptance of atypical combinations) (e.g. Siyanova and Schmitt 2007, 2008).

Processing and bilingualism (1):

1. Because language input is sequential and fleeting:
   i. processing makes use of prediction across multiple levels of representation (e.g. Hagoort 2009; Pickering & Garrod 2007, 2013);
   ii. learning must occur while language material is available, online. Language learning is therefore learning to process (e.g. Christiansen & Chater 2015).
2. Two factors influence learning during processing (e.g. Ramscar et al 2012):
   i. Positive evidence (reinforcement of successful prediction)
   ii. Negative evidence ('detuning' through prediction error)

Processing and bilingualism (2):

3. Background rates of non-informative co-occurences far outstrip informative ones; learning is therefore driven by discrimination (or prediction error)(e.g. Baayen 2010).
4. Bilinguals (by virtue of first language primings/ associations which are activated during processing) have a far higher background rate, and therefore greater uncertainty. More diffuse distribution of error diminishes prediction and learning from expectation violations.

Negative evidence in L1 (e.g. Ambridge 2013, Boyd & Goldberg 2011, Goldberg 2006, Ramscar et al 2012):

In first language development, negative evidence solves Baker's Paradox (how children learn that certain patterns are ungrammatical/ dispreferred rather than simply not yet heard):

how did you unsqueezed it . . .

I beed a good typewriter . . .

let's get brooming . . .

these flowers are sneezing me . . .

(Stilwell-Peccei, 2006)

Questions

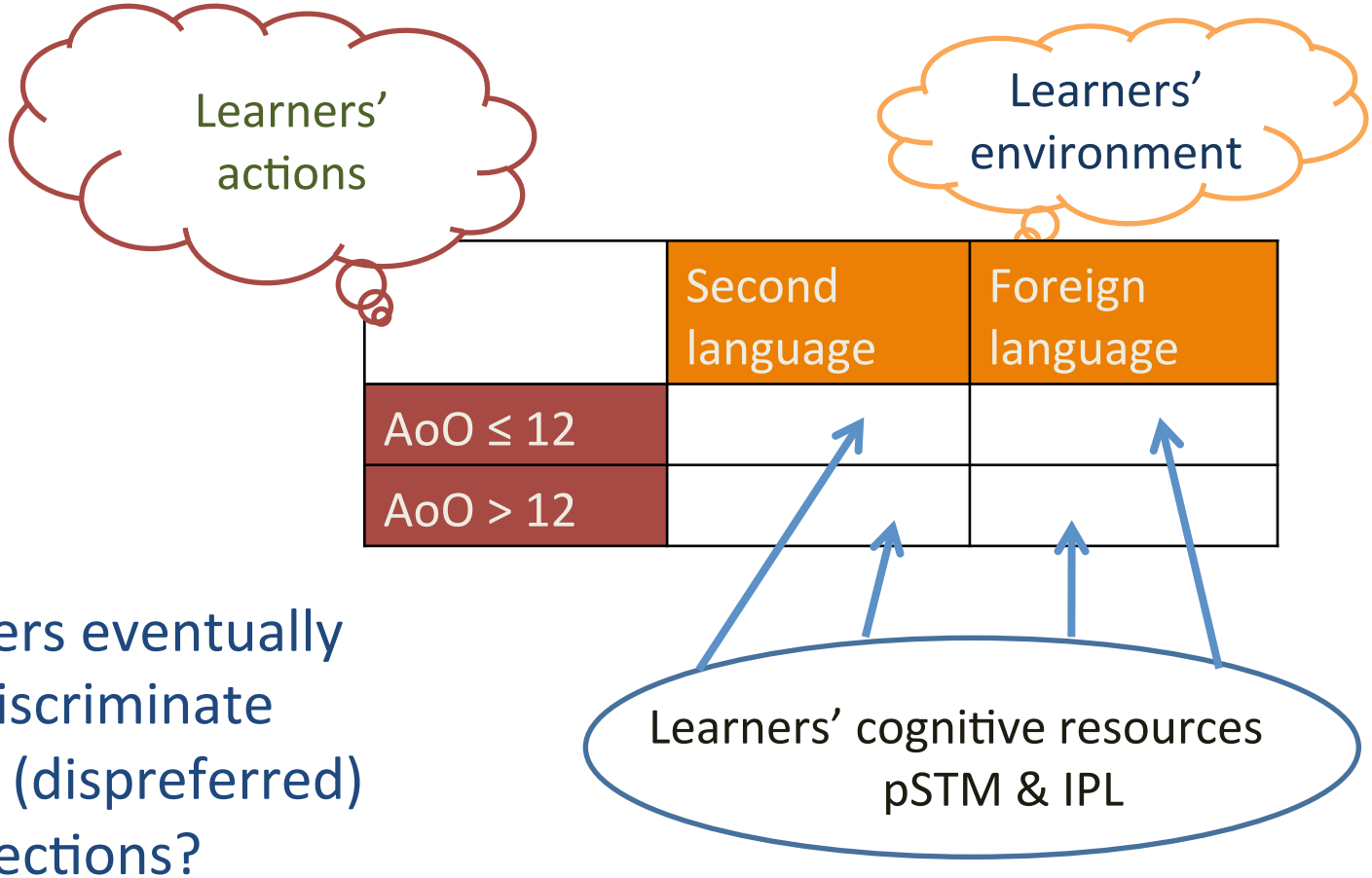
- Given massive amounts of exposure, can L2 users eventually learn to discriminate unnatural (dispreferred) lexical selections? If so, is there any evidence that they have accrued this knowledge in the same manner as native speakers?
  - Individual differences (ATI pSTM & incidental learning)
  - **Study 1 & 2**

- Does positive encoding lead to "appropriate associations"?
  - Explanatory power of absolute and relative frequencies
  - **Study 2 & in prep**

Bolibaugh (2014), Foster, Bolibaugh & Kotula (2014)

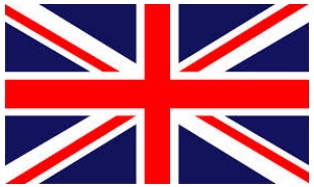
STUDY 1

"an alternative way of trying to understand processes that are hard or impossible to observe is to infer them from the way individual differences variables interact with linguistic & contextual variables" (DeKeyser 2012: 289).



Can L2 users eventually learn to discriminate unnatural (dispreferred) lexical selections?

Participants



≥ C1 CEFR

L1

L2

STUDY 1	STUDY 2
N = 39 Living in the UK > 12 yrs (*M*= 37.49, *SD*=18.46) Age of onset 1-35 years (*M*=18.85, *SD*=9.91) Age at testing (*M*=56.32, *SD*=16.61)	***N*** = 33 Living in UK > 10 yrs (*M*= 12.27, *SD*=3.25) Age of onset > 18 years (*M*= 23.88, *SD*=5.50) Age at testing < 50 years (*M*=37.15, *SD*=5.65)
N = 40 Living in PL, > 12yrs L2 use (*M*=16.68, *SD*=4.95) Age of onset 5-30 years (*M*=12.72, *SD*=6.25) Age at testing (*M*=29.40, *SD*=8.96)	

NLS texts

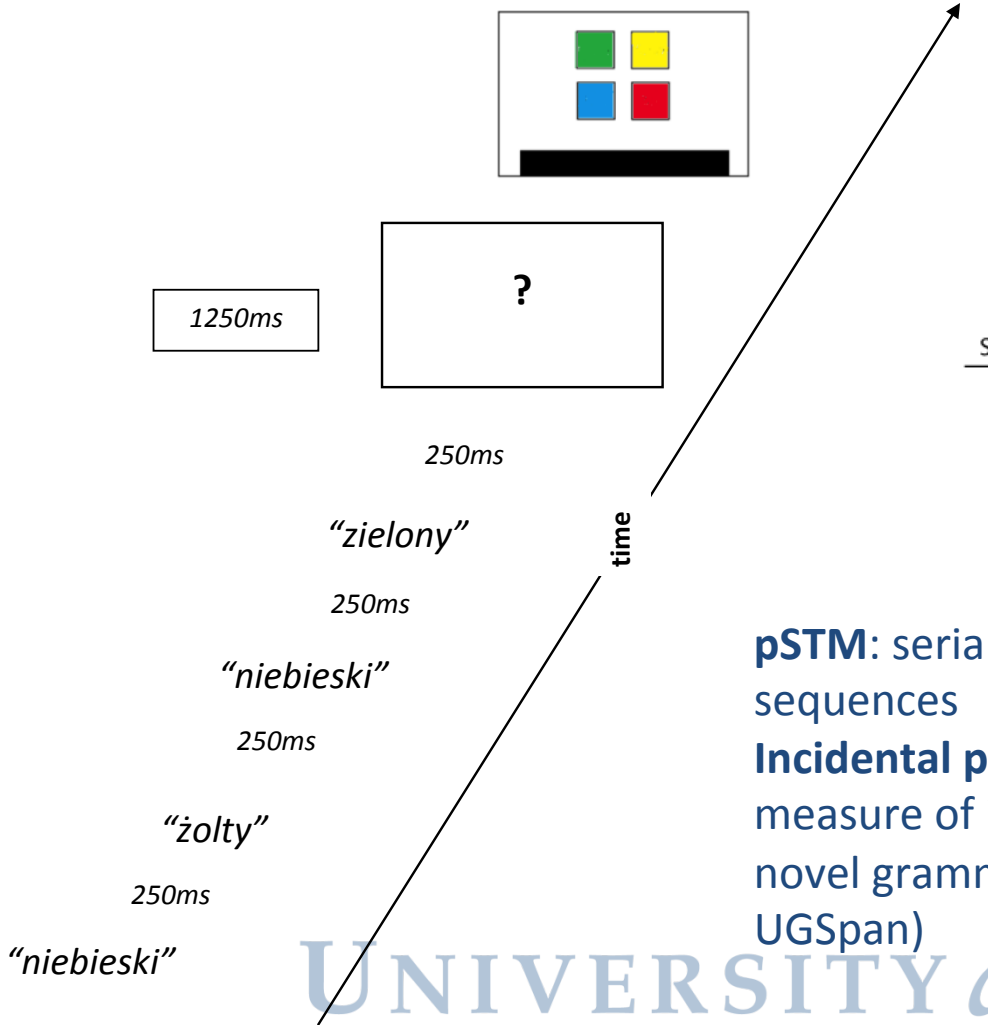
2 short texts (≈200 words each)

24 nonnative lexical selections

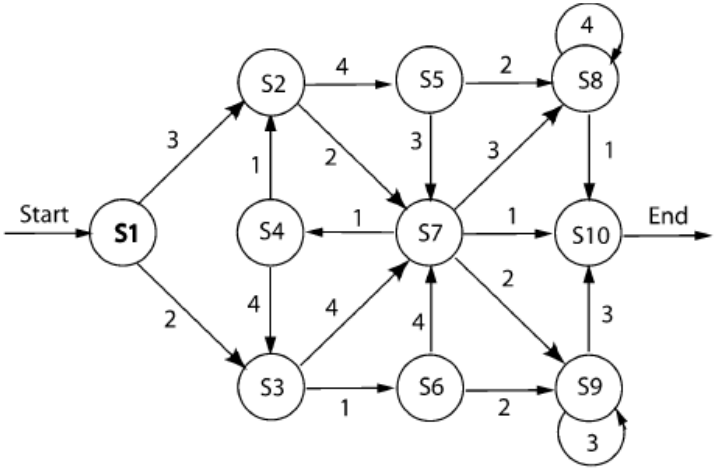
- collocation (*imagine an idea, get success*),
- derivational morphology (*to gunfight*),
- colligation (*the ball came up by floating, reply by a shrug*).

But one of the boys was cleverer than the rest and imagined a good idea.
He ran off, and soon came back with a big bucket of water.
He poured it into the hole fully. The ball came up by floating and they got it back…

Phonological short term memory (serial recall) & incidental pattern learning (Karpicke & Pisoni 2004)



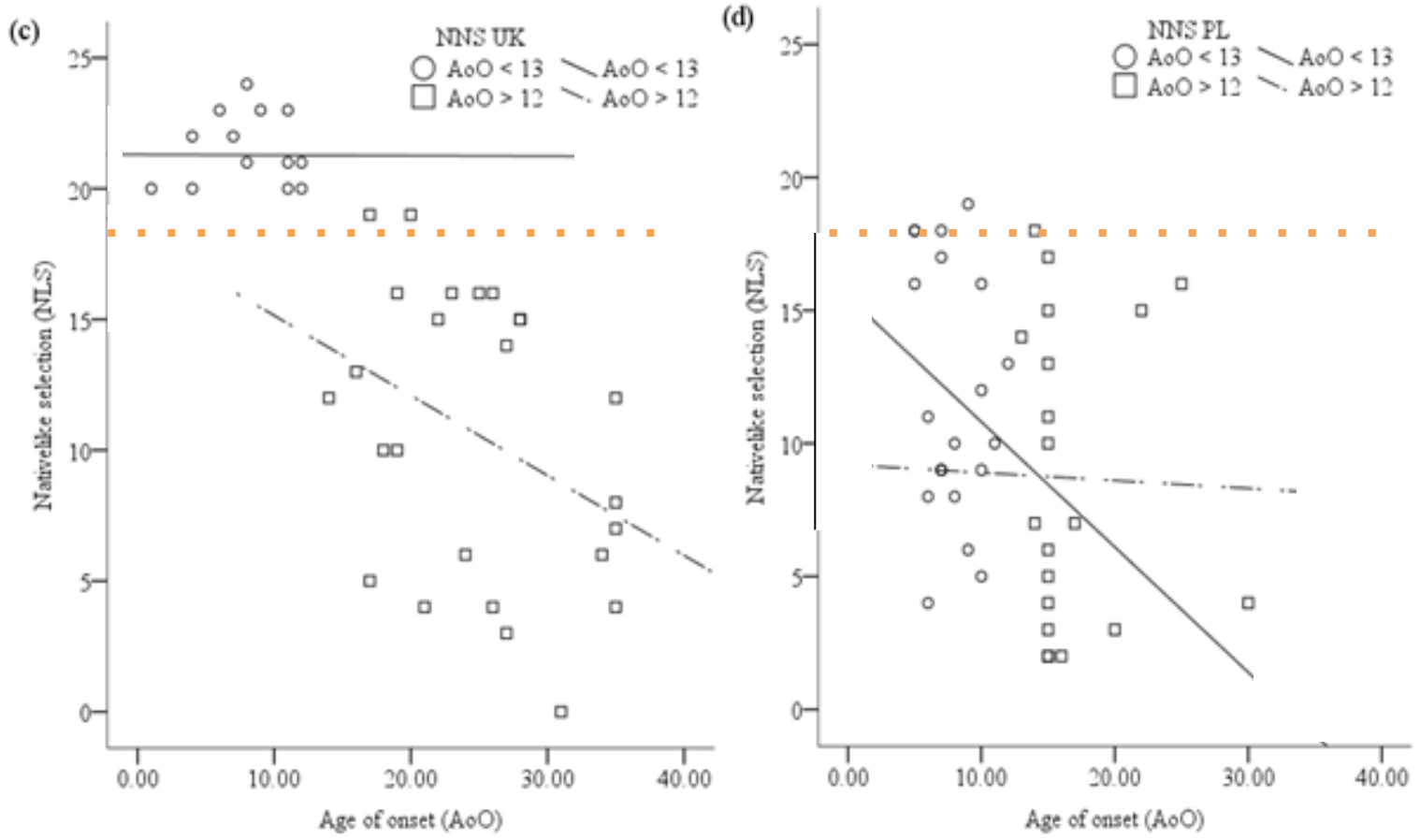
"blue yellow blue green"



pSTM: serial recall of ungrammatical sequences

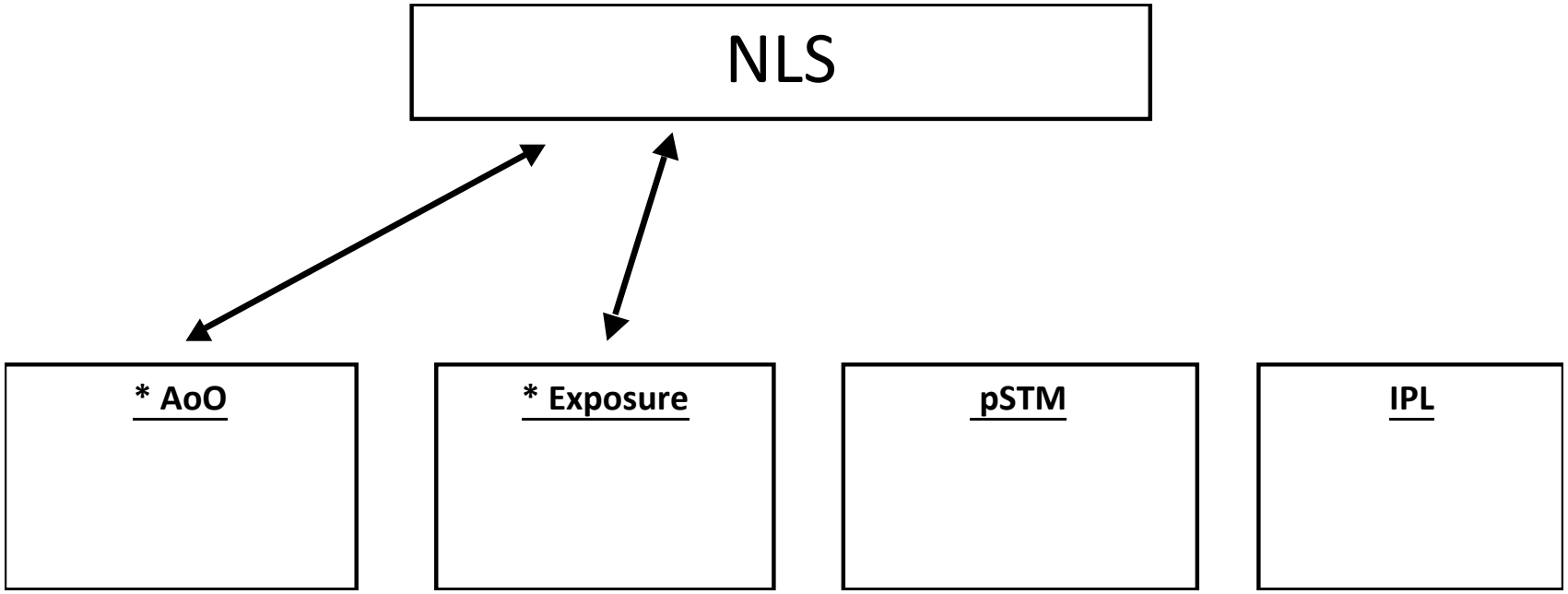
Incidental pattern learning (IPL): measure of improvement in memory for novel grammatical sequences (Gspan – UGSpan)

Findings: NS **> NNS_UK** > NNS_PL

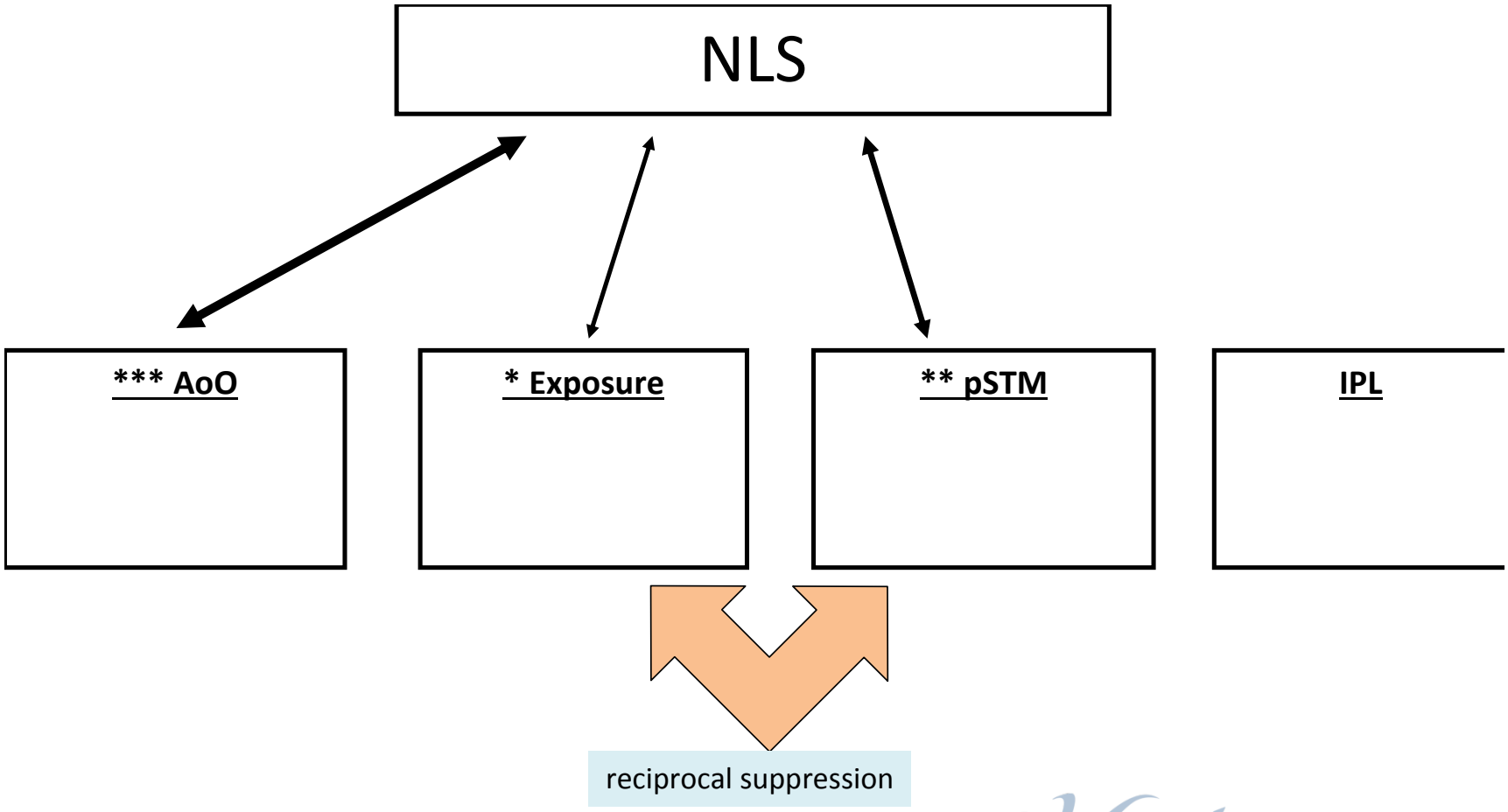


UNIVERSITY of York

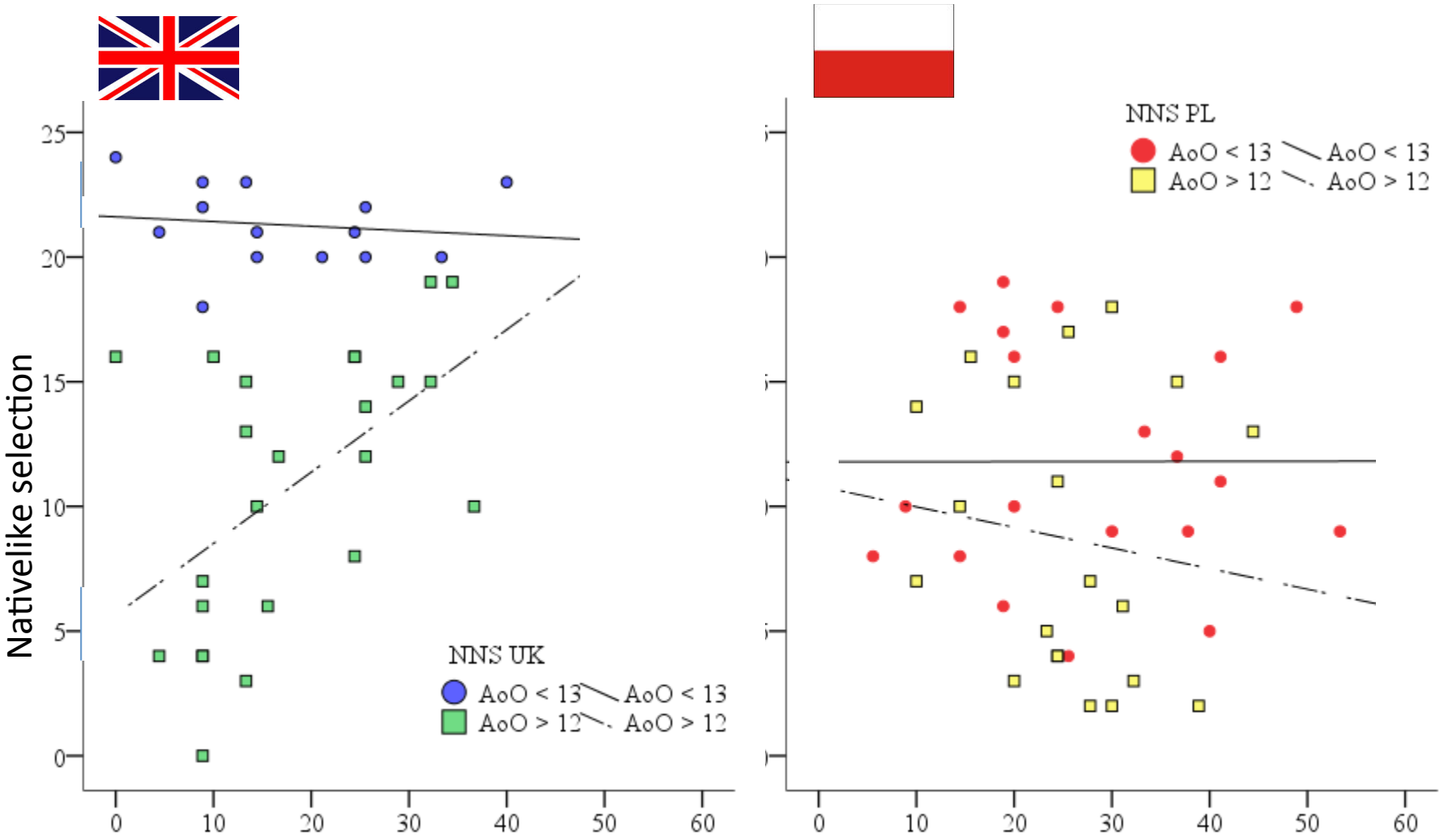
Findings NNS PL



Findings NNS UK



pSTM(study 1)



Phonological short term memory

Study 1: Findings

- ✓ Type of input Context significantly influences ability to discriminate atypical lexical selections (NNS UK *> NNS PL), and mediates effect of personal & ID variables
- ✓ Amount of input Long immersion increases discrimination ability but only brings two late starters up to NS range. Moderate effect in FL context.
- ✓ Evidence for incidental incremental learning Evidence of incidental learning in immersion: reciprocal suppression between length of exposure and pSTM in immersion. No main effect or interaction for pSTM in FL context.

- Age at testing?
  - Suppression might be a result of proxy effects;
- LoR as measure of exposure?
  - Depth of exposure unlikely to have remained constant over time

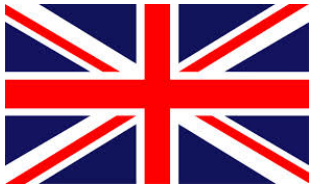
Bolibaugh & Foster (2013), Bolibaugh (2014)

STUDY 2

Partial replication & extension

- ✓ Do adult onset L2 users with long experience develop the ability to discriminate dispreferred lexical selections?
  - Is there evidence of interaction between depth of exposure & cognitive variables?

Participants



≥ C1 CEFR

L1 ***L2***

STUDY 1	STUDY 2
N = 39 Living in the UK > 12 yrs (*M*= 37.49, *SD*=18.46) Age of onset 1-35 years (*M*=18.85, *SD*=9.91) Age at testing (*M*=56.32, *SD*=16.61)	***N*** = 33 Living in UK > 10 yrs (*M*= 12.27, *SD*=3.25) Age of onset > 18 years (*M*= 23.88, *SD*=5.50) Age at testing < 50 years (*M*=37.15, *SD*=5.65)
N = 40 Living in PL, > 12yrs L2 use (*M*=16.68, *SD*=4.95) Age of onset 5-30 years (*M*=12.72, *SD*=6.25) Age at testing (*M*=29.40, *SD*=8.96)	

NLS task(Bolibaugh & Foster 2013, Bolibaugh 2014,)

. . . One of the films I've seen lastly . . .

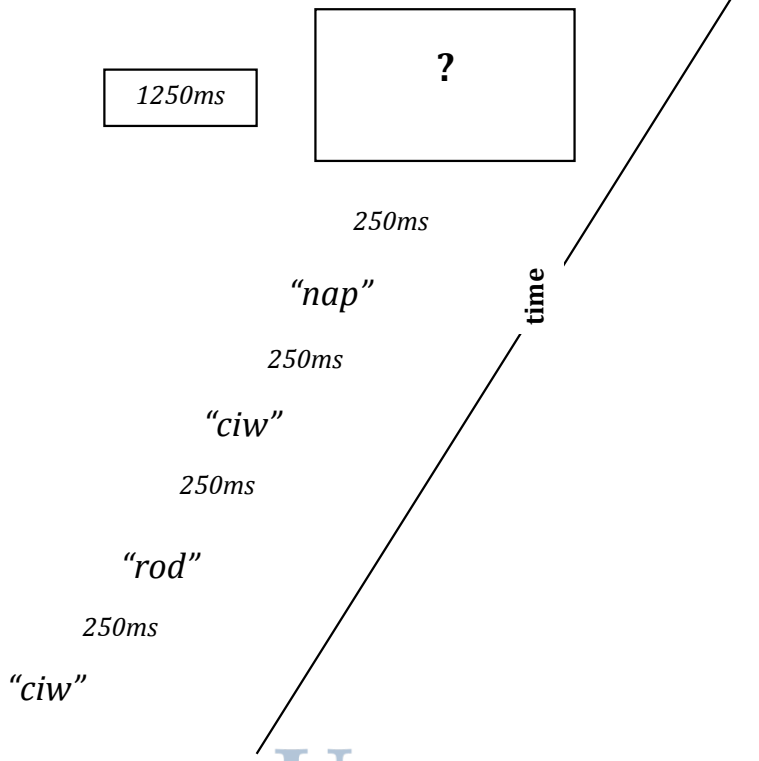
. . . by the director Mike Leigh. It's a modern time story. . .

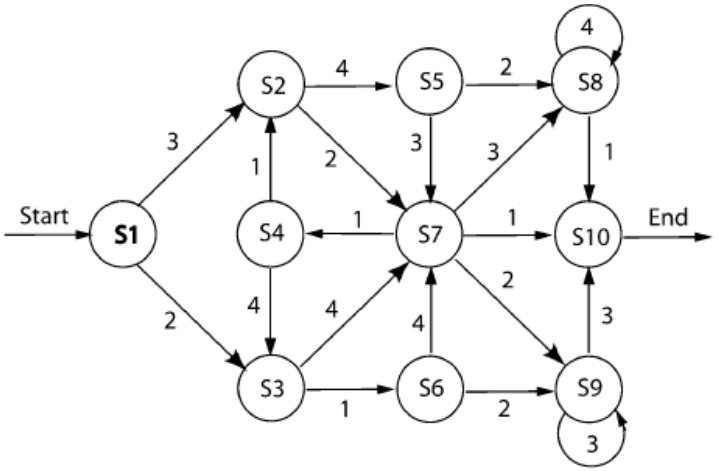
. . .It's a very strange film. I can't make the point of it. . .

. . . One girl has a father who is a heavy drunkard . . .

Phonological short term memory (serial recall) & incidental pattern learning (Karpicke & Pisoni 2004)

CVC nonwords



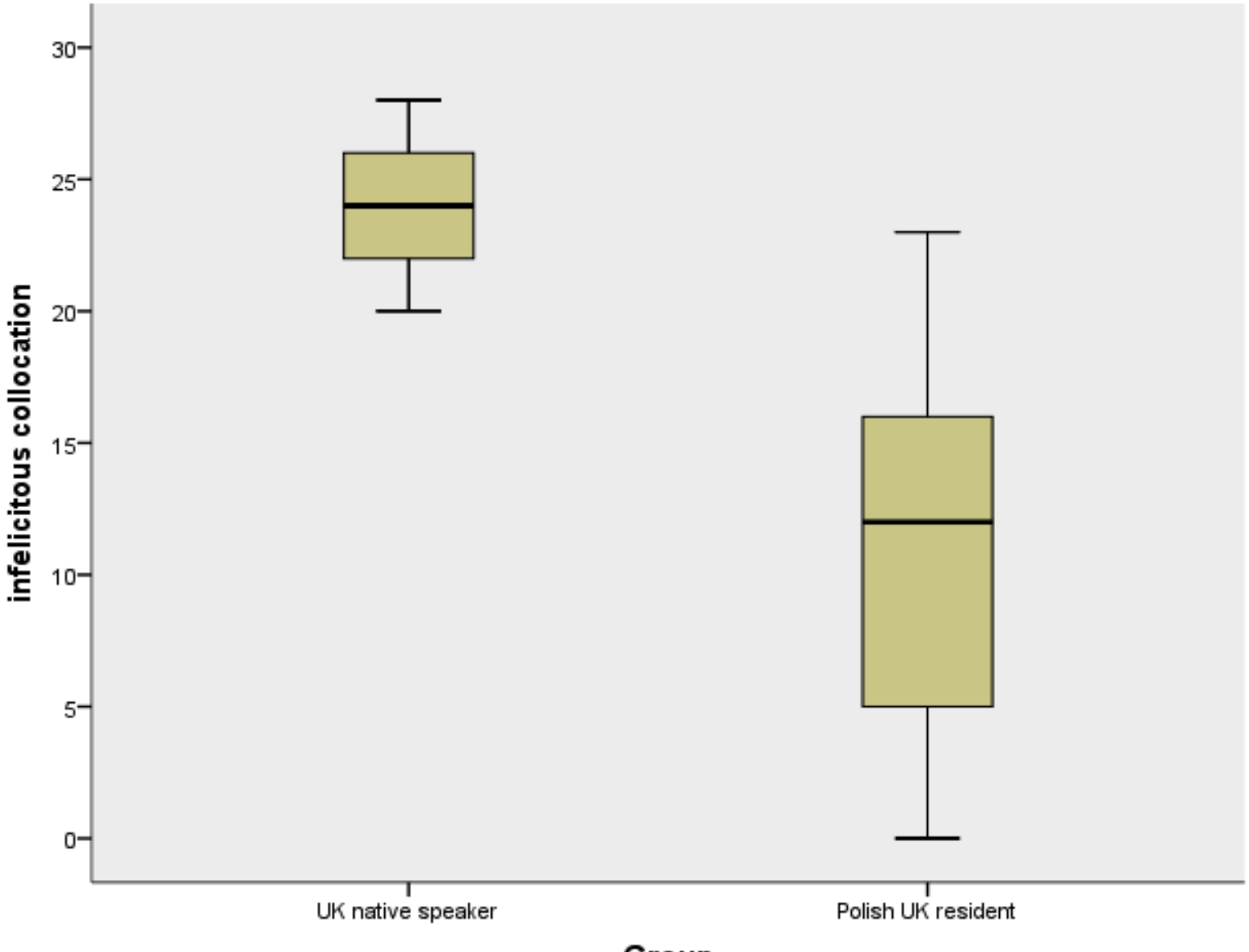


pSTM:
-serial recall ungrammatical sequences
-NWR repetition task

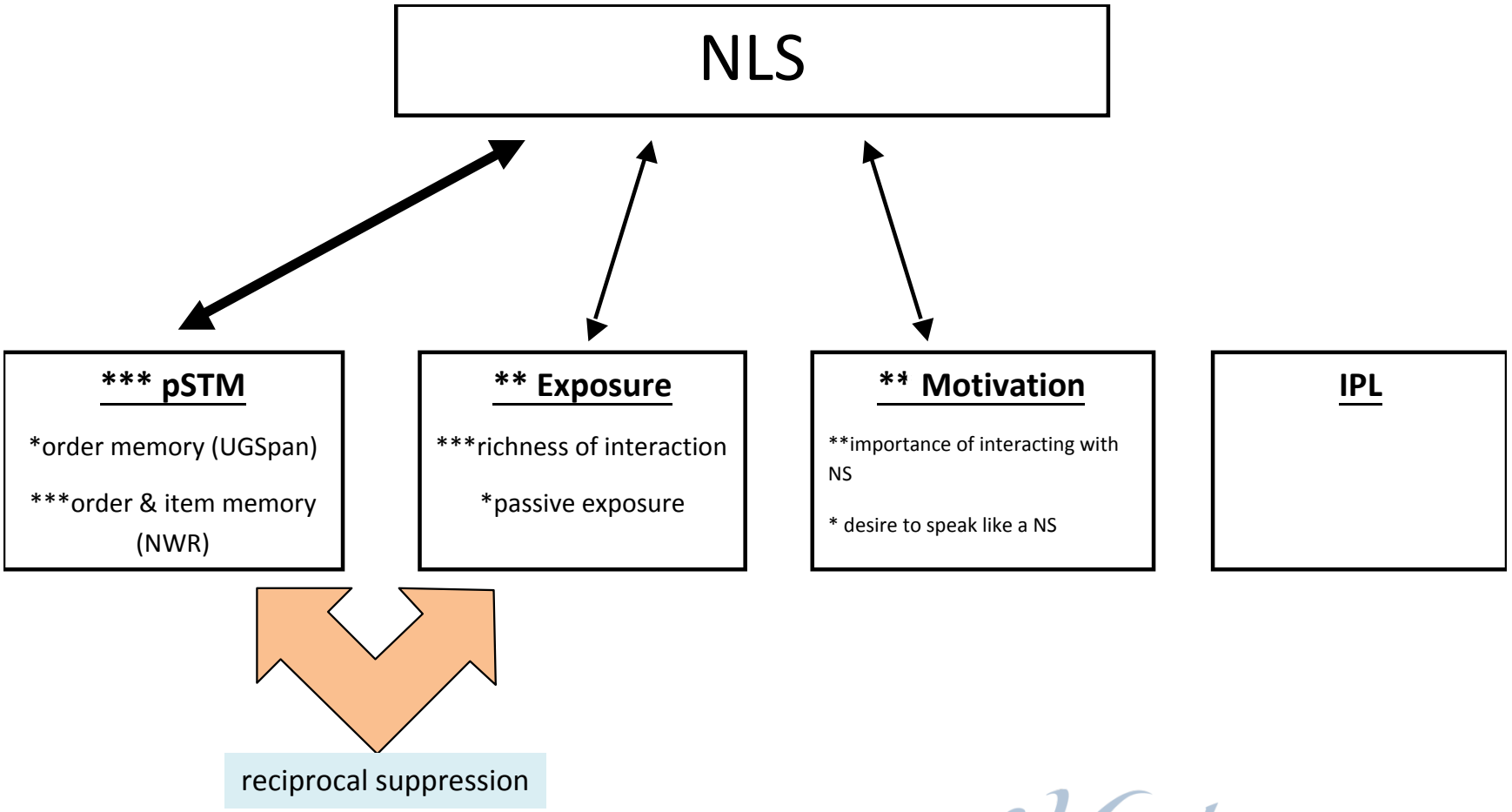
Incidental pattern learning (IPL):
measure of improvement in memory for novel grammatical sequences (Gspan – UGSpan)

UNIVERSITY of York

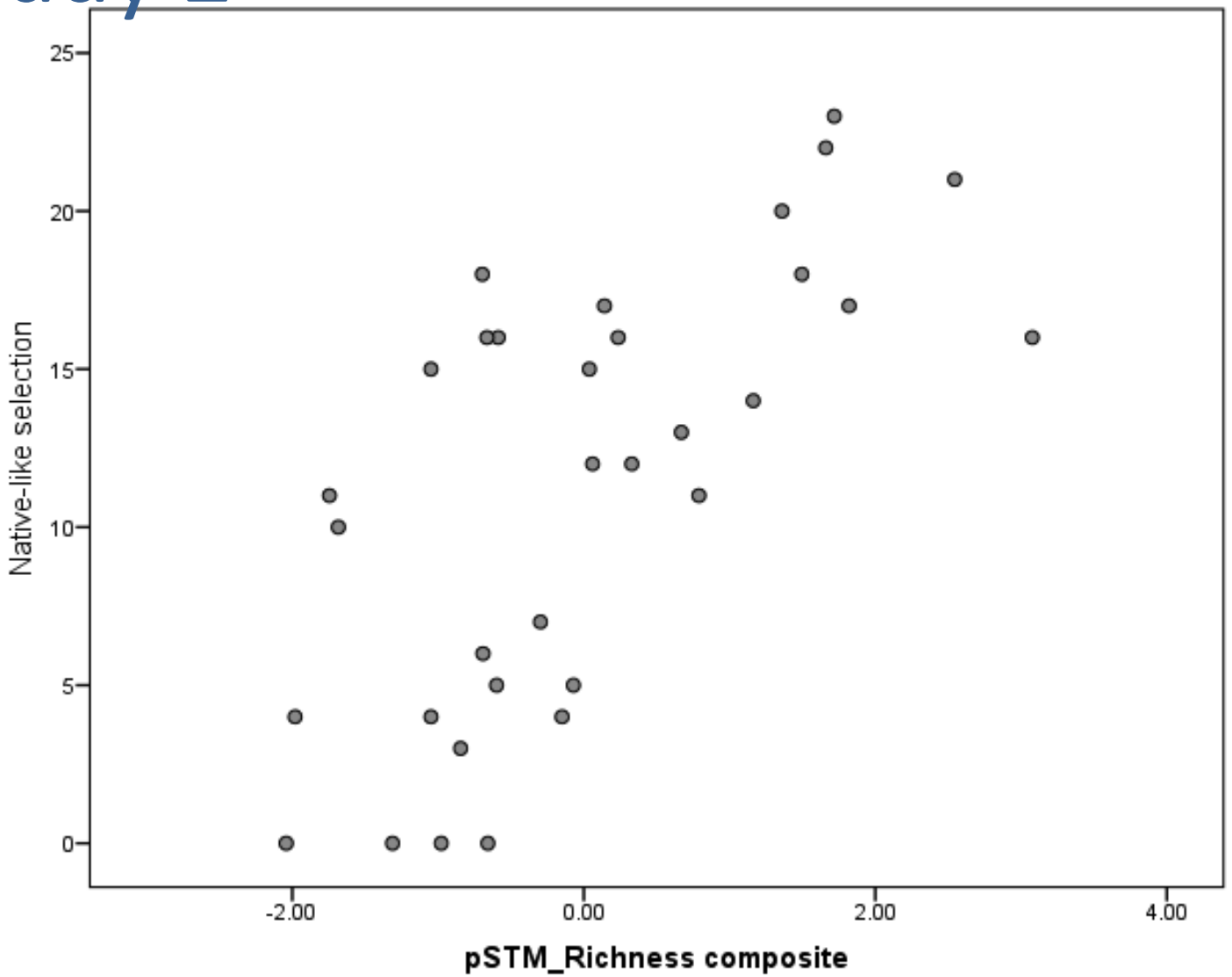
Study 2 findings (Bolibaugh & Foster 2013)



Study 2 findings



Study 2



R^2=.49, F(2,30) = 14.15, p =.001

Study 2: Findings

- ✓ Type & amount of input Richness of interaction associated with increased detection of dispreferred selections.

- ✓ Evidence for incidental incremental learning
  - ✓ reciprocal suppression between depth of interaction and pSTM confirmed.
  - ✓ 4 NNS who scored in NS range had very high pSTM & richness of interaction
- ✓ …still no effect of IPL

- ○ Distributional learning?
  - Unclear whether L2ers can go beyond item specifics

Is there evidence that L2ers are sensitive to negative evidence?

Bolibaugh (2014, in prep)

STUDY 3

What factors in the input account for variability in individual NNLS detection rates?

How do speakers learn not to use certain semantically sensical and syntactically reasonable formulations?

UNIVERSITY of York

The influence of absolute & relative frequencies

Likelihood of a novel/unattested collocation being pre-empted (e.g. marked, unacceptable, odd/not produced):

+ pre-emption: (LnFreq) of conventionalised alternative,

- uncertainty: (Hrel) of type frequency distribution.

Indirect negative evidence: (statistical) pre-emption/blocking (Goldberg 2006, Boyd & Goldberg 2011)

- (inflectional/derivational) morphology:

  [go + ed] pre-empted by *went*

  [good +er] *better*

  [steal + er] *thief*

- Collocation:

  [action].[V*] + a crime

  a)??do/make/perform

  b)commit

Verb argument constructions:

a) ?? She explained someone the story

b) She explained the story to someone

Constructional productivity:

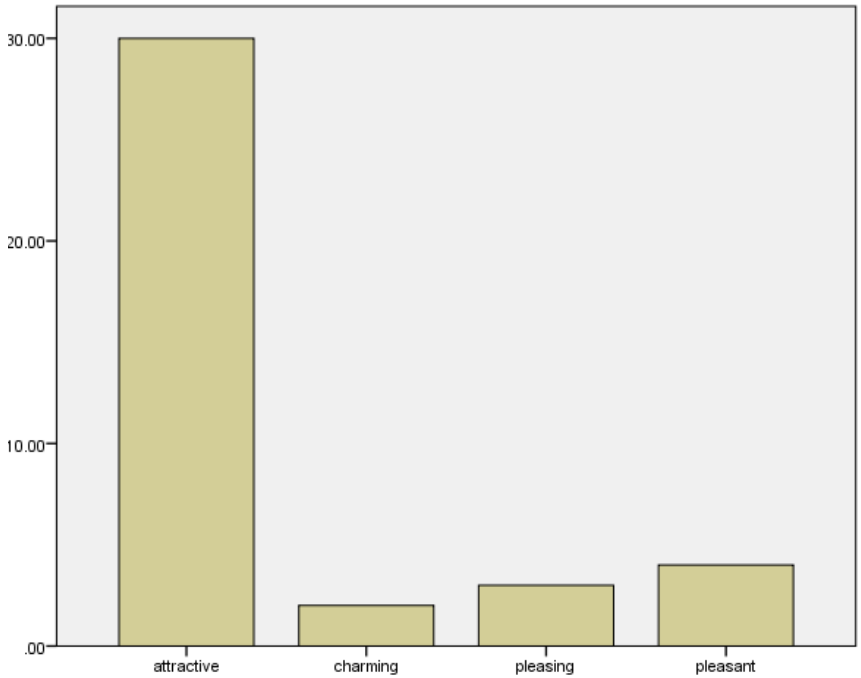
uncertainty as measured by H_{rel} (Baayen 2010, Gries 2012)

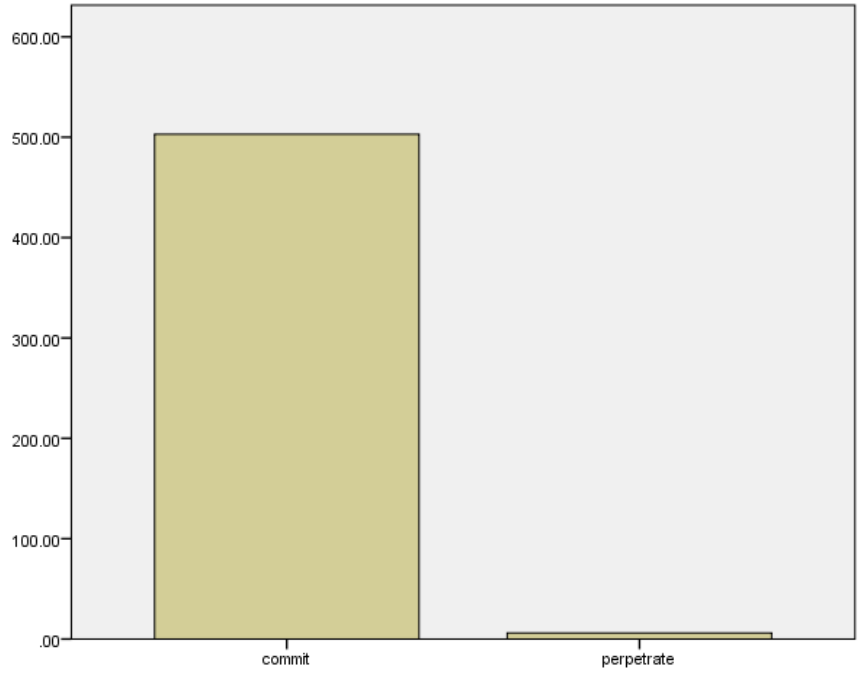
	Word 2	Word 3	Word 4
Word1	100	10	3

	Word b	Word c	Word d
Word a	400	370	120

measures the degree of mixed-upness (uncertainty) in the different states that a system can assume.

Pattern extendability





- **[POSITIVE ATTRIBUTE].Adj + *appearance***

[ACTION].V* + *crime/s*

H_{rel} = .57

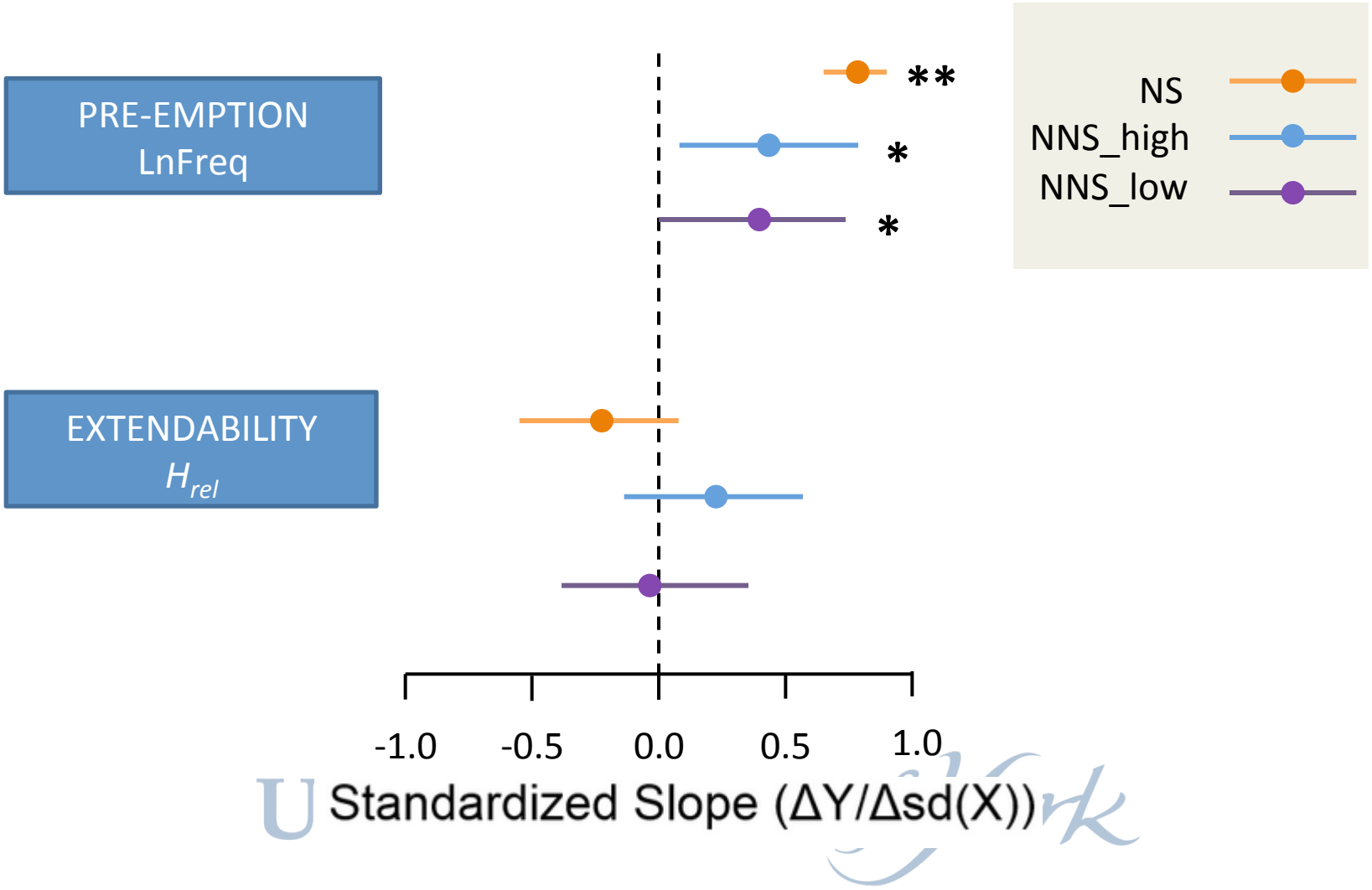
H_{rel} = .09

($H_{rel} = H(x)/ H_{max}$ $\sum_x p(x)\log_2 p(x)/ \log_2(x)$)

Item analysis
Extreme groups design

- Same participants as previous study
- Ranked according to pSTM_Exposure (index of mental corpus)
- Split into thirds (N=11).
- Item analysis for NS, NNS_high & NNS_low

Item detection rate (unacceptability) = $b_0 + b_1$ LnFreq $- b_2 H_{rel}$



	NS	NNS high	NNS low
Pre-emption	$\beta = .79$***	$\beta = .49$*	$\beta = .47$*
Extendability	$\beta = -.35$[a]	$\beta = .27$	$\beta = -.03$
R^2	.48**	.35*	.32*

*p <.05, **p <.01, ***p =.001, a) p =.08

Discussion

Results suggest that

- Second language speakers in the target language community do incidentally accrue information about idiomatic phrasings through use.
- Massive exposure does not automatically lead to nativelike intuitions (“appropriate associations’).
- L2ers are not sensitive to the distribution of error in the linguistic environment.

What's next?

Online study: magnitude of processing difficulty caused by dispreferred lexical selections (SPR/ eyetracking)

- 2 (typical/unattested) x 2 (+/- prototype frequency) x 2 (+/- pattern extendibility)

Questions?

Thanks! ☺

cylcia.bolibaugh@york.ac.uk

Collaborators:
Pauline Foster
Agnieszka Kotula

Partially funded by **ESRC** RES-000-22-2645

UNIVERSITY of York

References

- Ambridge, B. (2013). How Do Children Restrict Their Linguistic Generalizations? An (Un-) Grammaticality Judgment Study. *Cognitive science*, *37*(3), 508-543.
- Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, *122*(3), 292-305.
- Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. The Mental Lexicon, 5(3), 436-461.
- Bolibaugh, C. (2014) *Individual differences in second language nativelike selection ability: A usage based approach.* Unpublished PhD dissertation: University of Surrey.
- Bolibaugh, C., & Foster, P. (2013). Memory-based aptitude for nativelike selection: The role of phonological short-term memory. In Granena, G. and M. Long (eds.) *Sensitive periods, language aptitude, and ultimate L2 attainment*, *35*, 205.
- Boyd, J. K., & Goldberg, A. E. (2011). Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language*, *87*(1), 55-83.
- Christiansen, M. H., & Chater, N. (2015). The Now-or-Never Bottleneck: A Fundamental Constraint on Language. *Behavioral and Brain Sciences*, 1-52.

UNIVERSITY of York

- DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, *62*(s2), 189-200.

- Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*(2), 163-188.

- Ellis, N. C. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. *Vocabulary: Description, acquisition and pedagogy*, 122-139.

- *Ellis*, *N. C.* (*2001*). *Memory* for language. In P. Robinson (Ed.), Cognition and second language instruction (pp. 33–68). Cambridge: Cambridge University Press.

- Foster, P., Bolibaugh, C. & Kotula, A. (2014) Knowledge of nativelike selections in an L2: the influence of exposure, memory, age of onset and motivation in foreign language and immersion settings. Studies in Second Language Acquisition 36 (1), pp. 101-132

- Goldberg, A. E. (2006). Constructions at Work: The Nature of Generalization in Language. Oxford: Oxford University Press.

- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases. In A.P. Cowie (Ed.), *Phraseology: theory, analysis and applications.* Oxford: OUP.

- Gries, S. T. (2013). 50-something years of work on collocations: what is or should be next…. *International Journal of Corpus Linguistics*, *18*(1), 137-166.

- Gyllstad, H., & Wolter, B. (2015). Collocational processing in the light of a phraseological continuum model: Does semantic transparency matter?. *Language Learning*.

- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Psychology Press.

- Howarth, P. (1998). Phraseology and second language proficiency. *Applied linguistics*, *19*(1), 24-44.

- Karpicke, J. D., & Pisoni, D. B. (2004). Using immediate memory span to measure implicit learning. *Memory & Cognition*. 32(6), 956-964.

- Nesselhauf, N. (2004) *Collocations in a learner corpus*. Amsterdam: John Benjamins.

- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension?. *Trends in cognitive sciences*, *11*(3), 105-110.

- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329-347.

- Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of mouses in adult speech. *Language*, *89*(4), 760-793.
- Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word vs. one-word verbs. *International Review of Applied Linguistics in Language Teaching*, *45*(2), 119-139.
- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review/La Revue canadienne des langues vivantes*, *64*(3), 429-458.
- Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 776.
- Peccei, J. S. (2006). *Child language: a resource book for students*. Psychology Press.
- Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, *32*(4), 430-449.
- Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing. *Studies in Second Language Acquisition*, *35*(03), 451-482.
- Wray, A. (2002). *Formulaic language and the lexicon* (Vol. 5). Cambridge: Cambridge University Press.

Participants

	Polish in UK			Native Speakers		
	M	SD	N	M	SD	N
Age at testing (Age)	37.15	5.65	33	37.06	12.57	30
Years of Schooling (YrsSch)	15.82	2.04	31	16.22	2.77	30
Age of Onset (AoO)	23.88	5.50	33			
Age of Arrival (AoA)	25.18	5.22	33			
Length of Residence (LoR)	12.27	3.25	33			
Years Studying Eng before Arrival (YrsEng)	1.12	2.01	33			
Gender	24 F	9M		17F	13M	