

# Missing Generalizations: A Supervised Machine Learning Approach to L2 Written Production

**Daniel Wiechmann**

Amsterdam Center for Language and  
Communication  
University of Amsterdam  
d.wiechmann@uva.nl

**Elma Kerz**

Department of English Linguistics  
RWTH Aachen University  
kerz@anglistik.rwth-  
aachen.de

## Abstract

Recent years have witnessed a growing interest in *usage-based* models of language, which characterize linguistic knowledge in terms of emerging generalizations derived from experience with language via processes of similarity-based distributional analysis and analogical reasoning. Language learning then involves building the *right* generalizations, i.e. the recognition and recreation of the statistical regularities underlying the target language. Focusing on the domain of relativization, this study examines to what extent the generalizations of advanced second language learners pertaining to the usage of complex constructions differ from those of experts in written production. We approach this question through supervised machine learning employing as a primary modeling tool random forests with conditional inference trees as base learners.

## 1 Introduction

One of the central questions in second (L2) language learning is how L2 learners construct a new language system on the basis of only limited exposure to the target language. While formalist (*generative, syntax-based*) approaches have emphasized the reliance on innate mechanisms and principles, functionalist (*emergentist, usage-based* (UB)) approaches have highlighted processes of bottom-up induction of grammatical knowledge from input by way of complex automatic distributional analyses of perceived utterances at many grain-sizes (cf. Harrington, 2010 for an overview). The capacity to detect statistical regularities in the perceived input and to exploit these for purposes of building up more abstract generalizations is at work not only in

earlier stages of language acquisition, but remains throughout life (cf. Farmer, Fine & Jaeger, 2011), and is operative not only in the acquisition of L1 but also L2 (see, MacWhinney, 2013 for an overview). Grammatical knowledge then emerges through iterative categorization, in which the categories formed by grouping together similar exemplars at one level form the input of subsequent categorization processes at the next higher level of organization. In this view, language learning involves the task of identifying those variables that are involved in defining the generalizations that characterize conventional language use. In earlier stages of development, learners are found to establish generalizations along easily detectable, salient variables (MacWhinney, 2008). With growing experience, learners detect additional defining features and relationships among features and continue to refine their knowledge, resulting in their own productions become more and more target-like.

The resulting knowledge is likely to comprise both stored exemplars as well as generalizations derived through processes of analogical reasoning (see, e.g., Tomasello, 2003; Daelemans & van den Bosch, 2005; Goldberg, 2006; Ellis & Larsen-Freeman, 2009). At present, there is no general consensus as to what form the resulting knowledge takes and to what extent (if any) human linguistic knowledge is characterized by representational redundancy (cf. Wiechmann, Kerz, Snider & Jaeger, 2013 for a recent overview). Theoretical constructs to capture the units of linguistic regularity resulting from such processes of inductive learning include *local associations* and *memorized chunks* (Ellis, 2002), *computational routines* (O'Grady,

2005), and *constructions* (Tomasello, 2003; Goldberg, 2006; Langacker, 2008). In this paper, we assume the latter and follow a *UB constructionist* approach, in which all linguistic knowledge is characterized in terms of pairings of form and meaning, so called *constructions*. In this view, language learning concerns the emergence of symbolic units from the intricate interplay between “the memories of all the utterances a learner’s entire history of language use and the frequency-biased abstraction of regularities within them” (Ellis and Freeman, 2009:92). The emerging constructional patterns assume various degrees of abstraction and internal complexity and range from morphemes, to words and idiomatic expressions, to partially schematic (Kay and Fillmore, 1999) to fully schematic constructional patterns, such as clause-level argument structure constructions (Goldberg, 2006). Constructionist accounts are thus committed to the belief that “[a]n adequate model of human language processing must allow for a heterogeneous store of elementary units, ranging from single words, and basic combinatory rules, to multiword constructions with various open slots and complete sentences” (Beekhuizen, Bod & Zuidema (2013:267).

This study investigates knowledge about patterns at the sentential level, specifically knowledge about complex constructions involving relative clauses (henceforth RCs). RCs have played a pivotal role in the development of modern psycholinguistic theorizing and a lot of attention has been devoted to studying their acquisition and online processing (cf. Sheldon, 1974; Goodluck and Tavakolian, 1982; Diessel, 2004; Rohde, Levy & Kehler, 2011; Levy & Gibson, 2013; *inter alia*). In the domain of first language acquisition, UB constructionist accounts have portrayed the development of relative constructions types in terms of *clause expansion*, i.e. in terms of gradual transformations of simple (non-embedded) sentences into multiple-clause units (cf. Diessel & Tomasello, 2000; Tomasello, 2003; Diessel, 2004). In the domain of L2 learning, research on relativization has generally focused on assessing the degree to which L2 learning reflects the developmental pathways of L1 learning (Gass, 1979; Doughty, 1991; Abdomanafi & Rezaee, 2012). Largely based on comprehension tasks, these studies investigated if the learner proficiency in RCs decreases at lower positions of the accessibility hierarchy (Keenan and Comrie, 1977) and/or investigated related

proposals revolving around the internal syntax of relative clauses (e.g. the *Non-Interruption Hypothesis*, Slobin, 1973; the *Parallel Function Hypothesis*, Sheldon, 1974, or the *Perceptual Difficulty Hypothesis*, Kuno, 1975). This research has primarily addressed questions targeted at beginning and/or intermediate stages of L2 development of RCs. In recent years, there has been an increased interest in advanced stages L2 learning and harder to detect aspects of linguistic knowledge, which has resulted in a shift towards written production as “[...] in writing, rather than in speaking, the learner can [...] better show what he or she is capable of doing in and with L2 because writing allows far more reflection and is therefore usually somewhat more complex linguistically than speaking” (Verspoor, Schmid and Xu (2010:239). A growing availability of learner corpora of advanced L2 written productions gave rise to a number of studies whose main aim was to reveal factors of “foreign-soundingness even in the absence of downright errors” (Granger 2004:132). It was shown that - irrespective of their L1 background - advanced L2 learners face similar challenges on their way to near-native proficiency (DeKeyser, 2005; Wiechmann & Kerz, 2014) in connection with (a) a lack of register awareness and (b) an incomplete understanding of the complex probabilistic regularities underlying optional linguistic phenomena, which typically includes the integration of generalizations from various levels of organization (lexical, structural, discourse-pragmatic, etc.).

Focusing on advanced L2 learners’ written productions, the present study sets out to investigate a complex domain of grammar, viz. relativization. Specifically, we seek to understand the conditions in which experts prefer a reduced, non-finite RC over a more explicit, finite RC. The examples in (1) to (4) - taken from our expert data - illustrate the target structures. The modified nominal in the MC is referred to as the *head* of the RC.

- (1) The [<sub>head</sub> results] ] [<sub>RC</sub> that/which are shown in Tables IV and V] add to the picture [...]
- (2) The [<sub>head</sub> results] [<sub>RC</sub> shown in Tables IV and V] add to the picture [...]
- (3) The [<sub>head</sub> factors] ] [<sub>RC</sub> that/which are contributing to the natural destruction of microbes] [...]

(4) The [<sub>head</sub> factors ] [<sub>RC</sub> contributing to the natural destruction of microbes] [...]

We focus on the register of academic writing as it is characterized by a very condensed style (cf. Biber and Gray, 2010), which invites the increased usage of non-finite RCs. Furthermore, highly specialized domains, such as academic writing, afford specific register-contingent constructions (Kerz & Wiechmann, accepted).

## 2 Data

The data were retrieved from a corpus of 20 term papers produced by German students of English linguistics at RWTH Aachen University in their second and third year of study ( $N_{\text{words}} \sim 80,000$ ) and a same-sized control expert corpus of 10 peer-reviewed articles appearing in various journals on language studies. Manual extraction of all subject RC gave rise to a set of roughly 1,500 data points, of which 713 instances were produced by learners and 793 were produced by experts. All instances were manually annotated with respect to eight variables that have been shown to affect the online processing of RC constructions (cf. Fox and Thompson, 1990; Wiechmann, 2010 for a comprehensive discussion).

Variable	Description	Values
GROUP	item sampled from which group	advanced learner / expert
ID	source text	10 sources expert writing, 20 sources advanced learners
FINITE.RC	finiteness of RC	finite / non-finite
EXT.SYN	modified nominal in the MC	SU, DO, PN (predicate nominal), lower
LENGTH.BIN	length of sentence in words	dichotomized around the mean
ADD.MOD	presence of additional modifier (AP or PP)	yes / no
HEAD.TYPE	morphosyntactic type of head	lexical, pronominal, proper name
DEFINITE.HEAD	definiteness of head noun	definite / indefinite
ANIMACY.HEAD	animacy of head noun	animate / inanimate
GENERIC.HEAD	contentfulness of head noun	generic / specific
FREQUENT.AC	element of 100 most frequent heads in register (COCA/BNC)	yes / no

Table 1: Variables used in data description

The variables in Table 1 concern features of (a) the overall sentence (e.g. which grammatical role in the main clause is being modified by way of an RC, how long is the overall sentence, etc.) and (b) features of the head of the RC (e.g. does it refer to an animate or inanimate referent, is the nominal definite or indefinite, etc.).

## 3 Method

To assess to what extent the learners have successfully captured the regularities underlying the target system, we fit classification models to each data set that were geared to discriminate between finite or non-finite RC constructions based on the distributional information about the variables listed in Table 1. If learners have indeed successfully induced the right generalizations, then the models should reveal similar structures for both experts and learners. As a primary modeling tool, we used a random forest (RF) technique utilizing conditional inference trees as base learners (for details, cf. Hothorn, Hornik, & Zeileis, 2006; Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008; Strobl, Hothorn, Zeileis, 2009). We focused on this ensemble method for its ability to (a) produce reliable estimates of variable importance in scenarios of correlating predictors (Belsley et al., 1980) – which are the norm rather than the exception for linguistic choice phenomena like the one investigated here –, (b) for its ability to avoid biases towards categorical variables that have more levels, and (c) for their ease of interpretability. The criterion for stopping of an individual tree’s growth was based on multiplicity Bonferroni adjusted p-values from permutation tests suggested in Strasser & Weber (1999). Recursion was stopped when a hypothesis of independence could not be rejected at  $\alpha = 0.05$ . We evaluated the RF model on the basis of classification accuracy via repeated random sub-sampling validation (100 iterations; random split: 70% training data – 30% test data) and compared its performance with a logistic regression model (GLM) including only main effects and a support vector machine (SVM) with an RBF kernel. Average classification accuracy for the expert data ranged from 69% for the GLM, to 70% for the RF technique to 72% for the SVM. The performance of identical models on the learner data was about 5% higher on average. To estimate the degree of heterogeneity of the RC productions that is due to individual author(s) and L2 learners respectively, we also fit

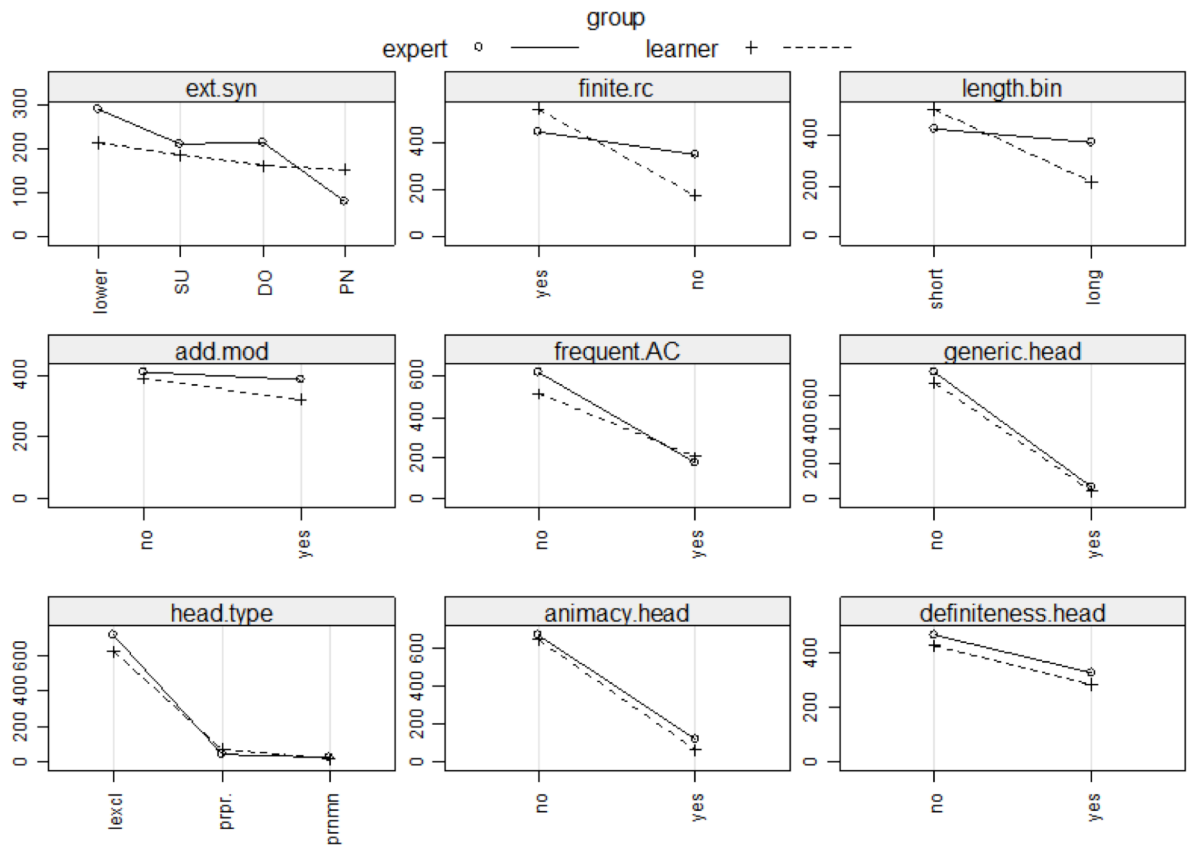


Figure 1: Distributions RC features: learners vs. experts

generalized linear mixed models (GLMM) to the data that in addition to the variables of interest also contained the variable ID (indicating the source text) as a random effect and investigated the adjustments to the intercept as an estimate of the degree of heterogeneity of the RC productions.

## 4 Results

Figure 1 presents an overview of the distributions of the descriptive variables in expert and learner productions.

### 4.1 Target-like productions

Figure 2 presents the results of a single conditional inference tree fit to all available data points from the expert set. In this model, the most important variable concerns the animacy of the head of the RC: in target-like productions, non-finite variants are more likely to be chosen when the modified nominal is inanimate (split at Node 1). Within the set of modifications of inanimate head nouns (Node 2), RCs non-finite variants are strongly preferred when the modified nominal functions as the subject of the dominating clause (Node 4). Within the subset of

non-subject modifications, the likelihood of an RC to be non-finite is greater when it is definite (Node 5). The model asserts additional structure with reference to the external syntax of the RC and the presence of an additional modifier to create a total of eleven partitions before tree growth is stopped. As individual trees are susceptible to small changes in the data, which typically leads to trees exhibiting high degrees of variability in their predictions, we checked the structure reported in Figure 2 against the relative variable importance derived from 500 trees with three variables randomly sampled as candidates at each node. Following Strobl, Malley and Tutz (2009), we considered variables to be non-important if their importance is negative, zero or has a small positive value that lies in the same range as the negative values. The RF model supports the important roles of all variables in the reported tree (relative importance in ascending order: FREQUENT.AC -0.002, HEAD.TYPE: 0.002, GENERIC.HEAD: 0.004, LENGTH: 0.005, ADD.MOD: 0.013, EXT.SYN: 0.013, DEFINITENESS.HEAD: 0.018, ANIMACY.HEAD: 0.036). We next estimated the variation that is due to individual stylistic differences in the ten texts that constitute our

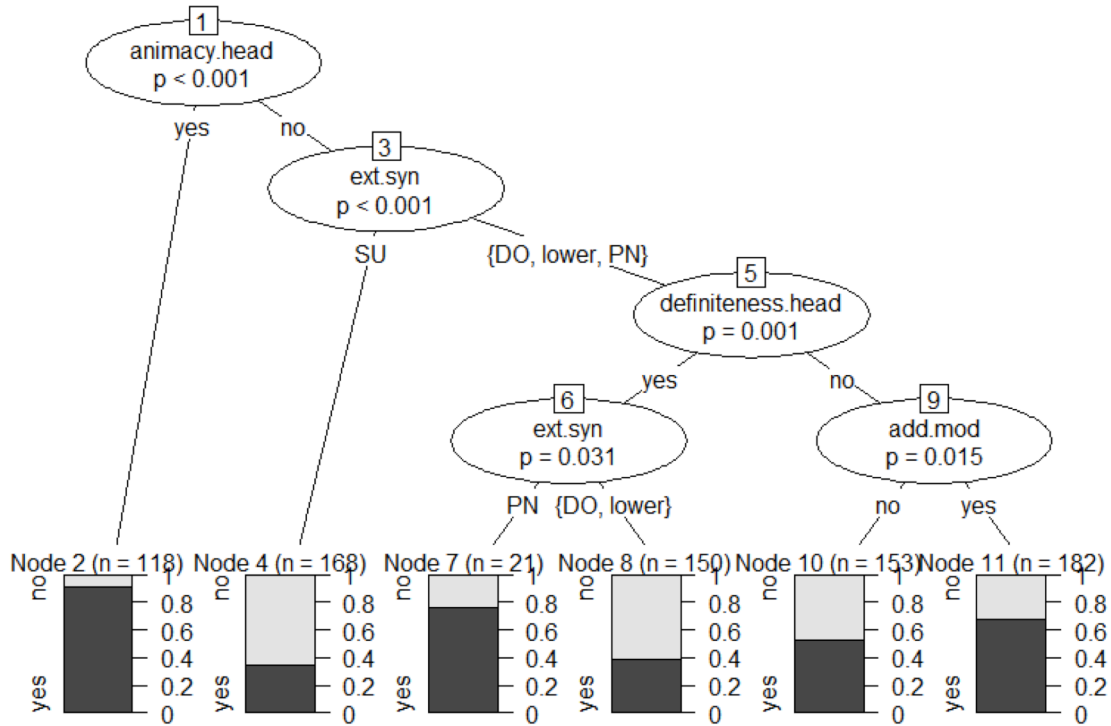


Figure 2: Conditional inference tree for expert data. Nodes contain Bonferroni-adjusted P-values ( $\alpha = 0.05$  as stopping criterion)

expert data using a GLLM that contained ID (source text) as a random effect. To avoid unnecessary model complexity, we excluded FREQUENT.AC, which was demonstrably unimportant for the constructional choice. As shown in Table 2, all effects were statistically significant at  $\alpha = 0.05$  (no 2-way nor 3-way interactions was significant at  $\alpha = 0.05$ ). The variability in the intercept between the texts in the expert corpus is negligible, suggesting that the relationships between the variables are rather robust in the target register (ID intercept variance = 0.07, SD = 0.26). Figure 3 shows the conditional modes of the random effect ID.

	Coef	SE	z	Pr(> z )
(Intercept)	0.31	0.21	1.45	0.15
ANIMATE.HEAD – no:yes	2.38	0.36	6.64	0.00
EXT.SYN – DO:lower	-0.12	0.20	-0.62	0.53
EXT.SYN – DO:PN	0.57	0.30	1.89	0.06
EXT.SYN – DO:SU	-0.48	0.22	-2.16	0.03
HEAD.TYPE – lex:pron	0.69	0.87	0.80	0.42
HEAD.TYPE – lex:name	0.91	0.38	2.38	0.02
LENGTH – long:short	-0.43	0.16	-2.68	0.01
ADD.MOD – no:yes	0.43	0.16	2.62	0.01
GENERIC.HEAD – no:yes	1.15	0.41	2.77	0.01
DEFINITE.HEAD – no:yes	-0.73	0.17	-4.24	0.00

Table 2: Generalized linear mixed logit model fit by the Laplace approximation (expert data)

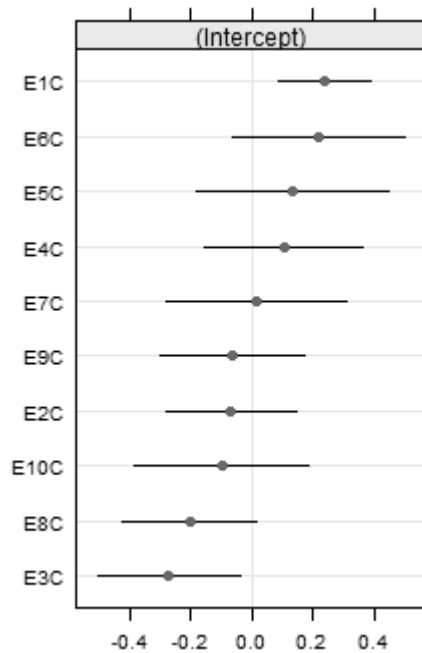


Figure 3: Conditional modes for the random effect ID in GLMM fit to expert data

## 4.2 Learner productions

We applied the exact same procedure to the learner data. We first present the results of a tree-based model fit to all exemplars in the learner

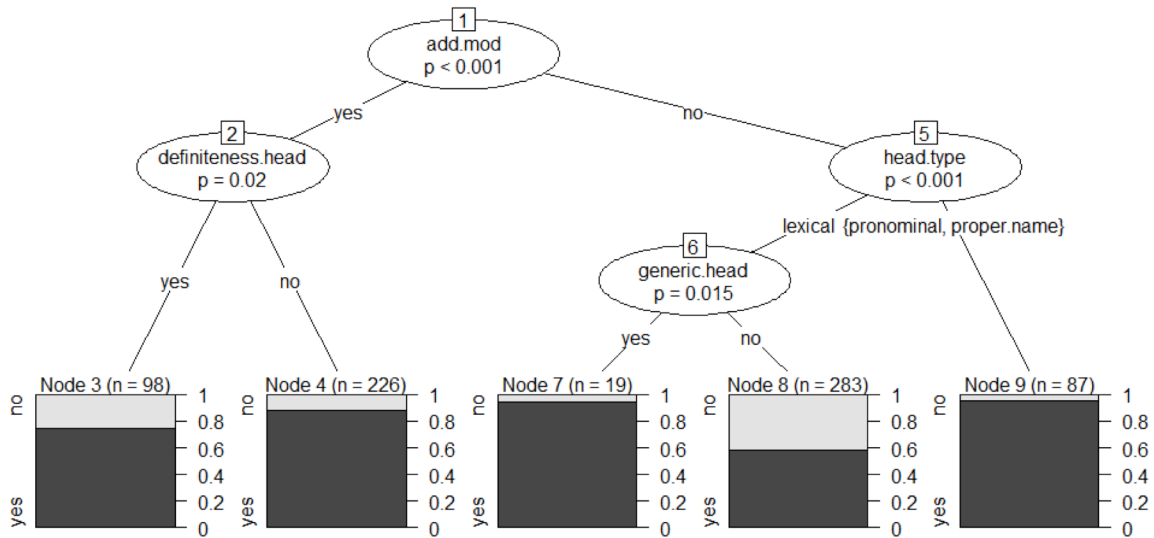


Figure 4: Conditional inference tree for learner data. Nodes contain Bonferroni-adjusted P-values (alpha = 0.05 as stopping criterion)

data (Figure 4). We found that the structure underlying the learner data is (a) simpler than the expert structure and also (b) different than the expert structure. At the top level, the data are split relative to whether or not there is an additional element to modify the head noun: the likelihood of a non-finite RC is slightly greater in the presence of an additional modifier and in particular with lexical heads that are not generic. The variable importance estimates derived from a model comprising 500 trees supported the importance of ADD.MOD, DEFINITENESS.HEAD, and HEAD.TYPE but not the importance of GENERIC.HEAD (relative importance in ascending order: GENERIC.HEAD = 0.002, ANIMACY.HEAD = 0.003, FREQUENT.AC = 0.003, LENGTH = 0.003, EXT.SYN = 0.004, DEFINITENESS.HEAD = 0.006, HEAD.TYPE = 0.006, ADD.MOD = 0.0205). The GLMM presented an overall similar picture supporting the importance of GENERIC.HEAD.

	Coef	SE	z	Pr(> z )
(Intercept)	0.92	0.36	2.52	0.01
ANIMATE.HEAD – no:yes	-0.32	0.38	-0.85	0.40
EXT.SYN – DO:lower	0.05	0.29	0.15	0.88
EXT.SYN – DO:PN	0.02	0.32	0.05	0.96
EXT.SYN – DO:SU	-0.25	0.30	-0.85	0.39
HEAD.TYPE – lex:pron	0.79	1.25	0.63	0.53
HEAD.TYPE – lex:name	3.70	0.77	4.81	0.00
LENGTH – long:short	-0.12	0.23	-0.52	0.60
ADD.MOD – no:yes	1.14	0.21	5.33	0.00
GENERIC.HEAD – no:yes	3.42	1.16	2.94	0.00
DEFINITE.HEAD – no:yes	-0.67	0.23	-2.94	0.00

Table 3: Generalized linear mixed logit model fit by the Laplace approximation (learner data)

Furthermore, the variability in the intercept between learners is a more pronounced than that of the experts (Figure 5).

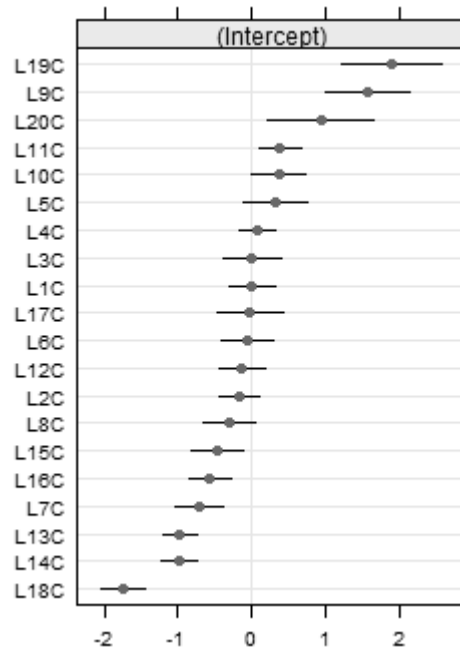


Figure 5: Conditional modes for the random effect ID in GLMM fit to learner data

## 5 Discussion

Our results indicated that advanced learners have clearly not yet built up the generalizations that characterize expert productions of non-finite RC constructions: firstly, the learners clearly underused non-finite variants of RCs relative to finite ones as evidenced by an observed ratio of

finite RC to non-finite RC of roughly 2:1 in learner language (compared to almost even proportions in expert language). As learners typically seek to maximize the transfer of knowledge from their L1 (MacWhinney, 2013), we assume that the underuse is at least partly due to the fact that there is no transferrable isomorphic translational equivalent to English nonfinite RCs in their L1 (German). However, this assessment clearly goes beyond the available evidence and falls outside the scope of this study. Secondly, our learners have derived generalizations that are less complex than those characterizing expert productions. Thirdly, they have assigned too much importance to some generalizations, e.g. the role of additional modifiers, and too little importance to others, e.g. animacy of the head noun and the external syntax of the RC. A linguistic analysis of relative constructions, which we will sketch only very briefly here, revealed that all variables to distinguish non-finite from finite subject relatives in expert language are semantically motivated. For example, in expert language non-finite RCs were strongly preferred in contexts where the RC modifies an inanimate, definite, lexical head that is the grammatical subject of the main clause as in (5).

- (5) The logic [used to resolve errors here] comes from the Cancellation/ Domination Lemma of Prince and Smolensky (1993:148) [...]

In such contexts the RC is almost invariably non-restrictive, i.e. its function is not to restrict the set of possible referents of the nominal, but rather to attribute a secondary predication to an already established discourse referent, while the main predication about that referent is encoded on the main clause (Wiechmann, 2010). The marginal adjustments to the intercept in the GLLM fit to the expert data suggested that the effects of these variables on the choice of RC are robust in the target register. In contrast, none of the constitutive features of this construction characterized non-finite RCs in learner language. The variable to distinguish the contrasted structural realizations of RCs in learner language most strongly was the presence of an additional modifier. An RC modifying a nominal that contains further pre- or post-modification was more likely to be realized in full finite form. Closer inspection of the data suggested that this preference does not reflect a semantic motivation

but rather reflects the tendency of language users to prefer explicit variants over reduced ones in contexts of greater complexity (Rohdenburg, 2003). Outside the context of semantically motivated constructions, expert language exhibited this preference as well, but its effect on the structural choice was noticeably less pronounced. We also found that the variability in the intercept is not very high suggesting that the generalizability of our findings is not threatened by the variability of the subjects' abilities to identify relevant generalizations. We found that about 80% of the learners formed a rather homogeneous group resulting in marginal adjustments to the intercept.

On a methodological note, we would like to briefly address two points: First, our approach to investigate (missing) generalizations does not speak to the issue of what exactly are the productive units in language and how exactly the operations of combinations are to be conceived of (for discussion cf. Bod 2009 and references therein) and does thus not constrain the computational realization of the statistical induction processes underlying language learning (cf. Clark, 2001; Klein and Manning, 2002; Zuidema, 2006; Bod & Smets, 2012; *inter alia*). In this paper, we were interested to what extent advanced L2 learners have succeeded in identifying generalizations pertaining to variables that figure in psycholinguistic accounts of sentence-level processing (e.g. animacy and definiteness of the head, type of embedding, etc.). Second, it was *not* the primary goal of our modeling to maximize predictive success. We address this point because we have also fit models based on much richer descriptions of the data (20+ variables) and some of these models reached levels of classification accuracy that exceeded that of the models reported here. However, we think that there are still good reasons to believe that their inclusion is actually detrimental to our attempts to understand the dynamics of language learning. To exemplify this: the variable 'voice of the RC' leads to an about 5% increase in classification accuracy of the expert model. However, its predictive value stems from the fact that it incorporates the effects of theoretically motivated variables thereby overshadowing their effects. Passive constructions tend to have inanimate subjects. As all RCs investigated here are subject relatives, this entails that the head of a passive RC tends to be inanimate. We find that 'voice of the RC' is more predictive than animacy of the head, but

the causal structure of the theory would have it that head animacy affects voice, rather than the other way round. With few exceptions, e.g. Baayen, Hendrix, and Ramscar (2013) on the reification of distributional effects, this general issue of predictors being robustly significant while lacking theoretical motivation has in our view not received the amount of attention it deserves. More generally, considerations like these motivate a shift towards the employment of causal models (cf. Pearl 2009).

## References

- Seyed Jalal Abdolmanafi and Abdolbaghi Rezaee. 2012. The difficulty hierarchy in the acquisition of English relative clauses. *International Journal of English and Education*, 1(2):170-179.
- Harald R. H. Baayen, Peter Hendrix, and Michael Ramscar. 2013. Sidestepping the combinatorial explosion: Towards a processing model based on discriminative learning. *Language and Speech*, 56(3): 329-347.
- Barend Beekhuizen, Rens Bod and Willem Zuidema. 2013. Three design principles of language: the search for parsimony and redundancy. *Language and Speech*, 56(3):265-290.
- David Belsley, Edwin E. Kuh, and Roy E. Welsch. 1982. *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons.
- Douglas Biber and Bethany Gray. 2010. Challenging stereotypes about academic writing: complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9:2-20.
- Rens Bod. 2009. Constructions at Work or at Rest? *Cognitive Linguistics*, 20(1), 129–134.
- Rens Bod and Margaux Smets, 2012. Empiricist Solutions to Nativist Problems using Tree-Substitution Grammars, *Proceedings Cognitive Models of Language Acquisition and Loss*, EACL 2012, Avignon, France, 10-18.
- Alexander Clark. 2001. Unsupervised induction of stochastic context-free grammars using distributional clustering. *Proceedings CoNLL 2001*, 105–112.
- Walter Daelemans, and Antal van den Bosch. 2005. *Memory-Based Language Processing*. Cambridge University Press, Cambridge, UK.
- Robert DeKeyser. 2005. What makes learning second language grammar difficult? A review of issues. *Language Learning*, 55:1-25.
- Holger Diessel. 2004. *The Acquisition of Complex Sentences*. Cambridge University Press, Cambridge.
- Holger Diessel and Michael Tomasello. 2000. The development of relative clauses in spontaneous child speech. *Cognitive Linguistics*, 11:131-151.
- Catherine Doughty. 1991. Second language instruction does make a difference: Evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition*, 13(4):431-469.
- Nick C. Ellis. 2002. Frequency effects in language processing. *Studies in Second Language Acquisition*, 24:143-188.
- Nick C. Ellis & D. Larsen-Freeman. 2009. Constructing a second language: analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning*, 59(1):93-128.
- Thomas Farmer, Alex B. Fine & T. Florian Jaeger. 2011. Implicit context-specific learning leads to rapid shifts in syntactic expectations. In L. Carlson, C. Hoelscher & T.F. Shipley (eds.), *Proceedings of the 33<sup>rd</sup> Annual Meeting of the Cognitive Science Society*, pp. 2055-2061. Cognitive Science Society, Austin, TX.
- Barbara Fox and Sandra A. Thompson. 1990. A discourse explanation of the grammar of relative clauses in English conversation. *Language*, 66:51-64.
- Susan Gass. 1979. Language transfer and universal grammatical relations. *Language Transfer*, 29:327-452.
- Adele Goldberg. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press, Oxford.
- Helen Goodluck and Susan Tavakolian. 1982. Competence and processing in children's grammar of relative clauses. *Cognition*, 11:1-27.
- Sylviane Granger. 2004. Computer learner corpus research: current status and future prospects. In U. Connor and T. Upton (eds.), *Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching*. John Benjamins, Amsterdam, 3-33.
- Michael Harrington. 2010. *Computational models of second language sentence processing*. In R. J. Kaplan (ed.), *Handbook of Applied Linguistics*, 2<sup>nd</sup> edition, pp. 189-204. Oxford University Press, Oxford, UK.
- Torsten Hothorn, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3): 651–674.
- Paul Kay and Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: The *What's X doing Y?* construction. *Language*, 75:1-33.



- Edward Keenan and Bernard Comrie. 1977. Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry*, 8:63-99.
- Elma Kerz and Daniel Wiechmann. accepted. Register-contingent entrenchment of constructional patterns: causal and concessive adverbial clauses in academic and newspaper writing. *Journal of English Linguistics*.
- Dan Klein and Chris Manning. 2002. A general constituent-context model for improved grammar induction. *Proceedings ACL 2002*, Philadelphia, 128–135.
- Susumu Kuno. 1975. The position of relative clauses and conjunctions. *Linguistic Inquiry*, 5(1):117-136.
- Roland W. Langacker. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford University Press, New York.
- Roger Levy and Edward Gibson. 2013. Surprisal, the PDC, and the primary locus of processing difficulty in relative clauses. *Frontiers in Language Sciences*, 4:229.
- Brian MacWhinney. 2008. A Unified Model. In N. Ellis & P. Robinson (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*, pp. 341-372. Lawrence Erlbaum Press, New York.
- Brian MacWhinney. 2013. The logic of the Unified Model. In S. Gass & A. Mackey (eds.), *Handbook of Second Language Acquisition*, pp. 211–227. Routledge, New York.
- William O’Grady. 2005. *Syntactic Carpentry: An Emergentist Approach to Syntax*. Erlbaum, Mahwah, NJ.
- Hannah Rohde, Roger Levy and A. Kehler. 2011. Anticipating explanations in relative clause processing. *Cognition*, 118(3):339-358.
- Günter Rohdenburg. 2003. Cognitive complexity and horror aequi as factors determining the use of interrogative clause linkers in English. In G. Rohdenburg and B. Mondorf (eds.), *Determinants of grammatical variation in English*, pp. 205-249. Mouton de Gruyter. Berlin.
- Dan I. Slobin. 1973. Cognitive prerequisites for the development of grammar. In Charles A. Ferguson and Dan Slobin (eds.), *Studies of Child Language Development*, pp. 175-208. Holt, Rinehart & Winston, New York.
- Amy Sheldon. 1974. The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, 13:272-281.
- Strasser H, Weber C (1999). On the Asymptotic Theory of Permutation Statistics. *Mathematical Methods of Statistics*, 8:220-250.
- Caroline Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin and Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics*, 9:307.
- Caroline Strobl, James Malley and Gerhard Tutz. 2009. An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychological Methods*, 14(4):323-348.
- Michael Tomasello. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press, Harvard.
- Marjolijn Verspoor, Monika S. Schmid & Xiaoyan Xu. 2012. A dynamic usage-based perspective on L2 writing. *Journal of Second Language Writing*, 21(3):239-263.
- Daniel Wiechmann. 2010. *Understanding Complex Constructions: A Quantitative Corpus-Linguistic Approach to the Processing of English Relative Clauses*. PhD Thesis. University of Jena.
- Daniel Wiechmann & Elma Kerz. 2014. Cue reliance in L2 written production. *Language Learning*.
- Daniel Wiechmann, Elma Kerz, Neal Snider & T. Florian Jaeger. 2013. Special issue: Parsimony and Redundancy in Models of Language. *Language and Speech*, 56(3).
- Willem Zuidema. 2006. What are the productive units of natural language grammar? A DOP approach to the automatic identification of constructions. *Proceedings CoNLL 2006*, 29–36.