Syntactic alternation research

Taking stock and some suggestions for the future

Stefan Th. Gries
University of California, Santa Barbara

Over the last 20 or so years, research on syntactic alternations has made great strides in both theoretical and methodological ways. On the theoretical side, much of the research on syntactic alternations was restricted to generative linguistics debating how near synonymous constructions differed slightly in meaning and/or how one (and which one) was derived from the other (transformationally). On the methodological side, much research consisted of monofactorial studies based on relatively simple text counts. By now, however, syntactic alternation research has become much more functional (in a broad sense of the term) and much more methodologically sophisticated: Much work is now motivated/interpreted psycholinguistically or in a broadly usage-based/cognitive linguistic framework and much work has now adopted a regression-based analytical strategy. These attractive developments notwithstanding, much remains to be done and, in this paper, I sketch some recent developments in (largely) separate alternation studies that I would like the field to adopt more broadly. These developments can be heuristically grouped into ones that have to do with (i) the statistical analysis of corpus-based and experimental alternation data, (ii) new predictors that explain typically unexplored aspects of variability in alternations.

1. Introduction

1.1 General introduction

During the past 50 or so years, alternations, in particular syntactic alternations, have been one of the most thoroughly researched kinds of topics. Their attraction to linguists can be explained by relating even a somewhat informal definition of the notion of alternation to linguistic methods and ideas. Specifically, if we define the notion of syntactic alternation as 'structurally and/or lexically different ways to say functionally very similar things', then it is immediately clear that alternations are a close relative to the notion of minimal pair in phonology, and it is immedi-

https://doi.org/10.1075/bjl.00001.gri
Belgian Journal of Linguistics Volume 31 (2017), pp. 8–29. issn 0774-5141 | e-issn 1569-9676

ately clear that alternations were an important issue in transformational-generative research (because they can constitute cases where different surface structures derive from the same deep structure). For the purposes of the present overview paper, I will distinguish two kinds of morphosyntactic alternations:

- word/constituent order alternations such as particle placement (*He picked up the book* vs. *He picked the book up*), the dative alternation (*He gave her the nuts* vs. *He gave the nuts to her*), the genitive alternation (*the speech of the President* vs. *the President's speech*), *to* vs. *ing* complementation (*he prefers to eat nuts* vs. *he prefers eating nuts*), and others;
- realization alternations such as *that* vs. Ø complementation (*I thought (that) there was an earthquake*) or reduced relative clauses (*the nut (that was) left on the table*), and others.

In much of the 20th century's transformational-generative work, alternations have both been viewed as not particularly relevant (in the sense that the choice for either structure is more of a performance rather than a competence issue) and as relevant in terms of the structure of the alternants and their derivational relations (in the sense that it needed to be explained how different surface structures may have been derived from the same deep structure); in keeping with much generative work of that time, the 'methodology' used for much of this work is based on judgments of grammaticality that were typically provided by the investigating linguists themselves.

In later functional (and then cognitive/usage-based) research beginning in the 1980s, linguists studied a variety of information-structural and cognitive/processing predictors of alternant choices such as the length of constituents and their definiteness as well as semantic and/or discourse-functional predictors such as animacy, givenness etc. of referents. In such older studies, linguists often used different kinds of (smaller) corpora, annotated relevant examples for the above-mentioned predictors, and did chi-squared tests or similar tests on predictors.

Finally, also during that time (although beginning a bit earlier with Labov's ground-breaking work), variationist sociolinguists explored the correlations of alternation choices with language-internal predictors just like those just mentioned as well as language-external predictors (such as sex, race/ethnicity, or socio-economic status of speakers); in most of these studies, researchers used a variant of binary logistic regression modeling called Varbrul, which given its idiosyncrasies (not allowing numeric predictors, making it hard to study interactions) is now thankfully slowly falling out of use.

Over the last 15 years or so, the above kinds of research have led to a now widespread adoption of approaches in which observational or experimental alternation data are analyzed with different kinds of (generalized) linear regression modeling; the most frequent case is probably that of binary logistic regression in which a variety of suspected alternation predictors and, crucially, also their interactions, are used to predict one of two constructional choices. This is a very good development for a variety of reasons: First, in ways not always recognized, even the statistically simplest monofactorial designs – e.g., chi-squared tests of whether one categorical predictors is correlated with a constructional choice – are, or can be recast as, regression models as when chi-squared tests are viewed as simple logistic or multinomial regressions. Second and as I have argued elsewhere (e.g., Gries 2011), sometimes studies that are portrayed as monofactorial are in fact multifactorial and, thus, require more powerful applications than simple chi-squared tests can offer. Specifically, since no alternation phenomenon is really monofactorial in nature, we need to be able to explore several predictors' roles and their interactions at the same time (see Gries and Deshors 2014, Section 3), or we need to be able to explore whether predictors' effects change over time, and the statistical framework of regression modeling allows to do just that. However, this welcome move towards more robust data than analysts' intuitions and more robust statistical analyses also comes with methodological challenges, some of which are well-known and some of which are maybe less well-known.

One particularly well-known challenge has to do with the assumption of many 'traditional' statistical tests that the data points are independent of each other, an assumption that typical data sets violate (because experimental subjects or speakers/writers of data in corpora provide more than one data point). For instance, using the alternation between the *will* and the *going-to* future, Gries (2016) shows that, leaving aside any and all linguistic or contextual predictors, one can predict about 90% of all constructional choices in the corpus used in that paper correctly just by (i) choosing the future that each speaker prefers in general and (ii) choosing the overall more frequent *will* when a speaker has no preference. State-of-the-art studies these days would use generalized linear mixed-effects modeling to capture speaker-/file-specific preferences with varying intercepts, i.e. the development of regression models that predict the outcomes of categorical dependent variables (here, *will* vs. *going to*) using intercepts and slopes for all data, but that can also take into consideration every speakers' overall baseline/preference by adjusting the intercept that all speakers share for each speaker (as needed); see Gries (2015).

Another well-known challenge to the independence-of-data-points assumption of simple regression modeling involves the observations that often lexical items have preferences for certain constructions (because of their semantic or information-structural characteristics), a notion that has been studied a lot in collostructional analysis of the kind proposed in Stefanowitsch and Gries (2003) and Gries and Stefanowitsch (2004). Again using future choice as an example, this time *will* vs. *shall* vs. *going to*, Gries (2016) shows that 87.7% of all future choices can be predicted correctly just on the basis of picking the most frequent future choice per lexical verb, e.g., predicting *will* for *go*, *give*, *come*, *do*, predicting *shall* for *receive*, *recite*, and predicting *going to* for *happen*, *say*, etc. This, too, would be addressed in state-of-the-art studies using verb-specific varying intercepts in a mixed-effects model.

A final and by now better-known challenge to the independence-of-data-points assumption are autocorrelation effects, e.g.

- when a previous use of an alternant X (the prime) makes another later realization of X (the target) less likely (because then two occurrences of X would be in too close succession), a principle, which is known as *horror aequi* and which could lead to a dispreference of *he was trying winning the game* because of the immediate succession of two *-ing* forms;
- when a previous use of alternant X (the prime) makes another realization of X next time around (the target) more likely (because of residual activation or implicit learning effects), a principle known as *persistence* or *priming* and which could lead to the use of a *will*-future by a speaker in a target construction when said speaker's or his interlocutor's last future choice, the prime, also was a *will*-future.

Gries (2016) shows that 80.9% of all future choices can be predicted correctly by choosing (i) the future the speaker used last time and (ii) *will* for every first future in a file. State-of-the-art studies of such alternations would therefore include a variable LASTCHOICE among the predictors, whose impact would ideally be moderated by the distance between the current choice under investigation and the last one, with maybe very short and long distances predicting inhibition and facilitation of another choice of X respectively.

The previous remarks have shown, if only briefly, that our regression models need a few more predictors than much work in the last century has considered, meaning our spreadsheets need a few more columns than one might think of at first. For every alternation choice, we need columns called FILE or SPEAKER to cover speaker-specific preferences, ITEM to cover lexically-specific preferences, and LASTCHOICE and/or DIST2LASTCHOICE to cover *horror aequi* as well as priming effects, followed by some version of generalized linear mixed-effects modeling … Unfortunately, this neither addresses all challenges nor exhausts all opportunities, which is why the remainder of this paper will be concerned with making a variety of suggestions for better and/or more comprehensive analyses of alternation phenomena. Specifically, Section 2 is concerned with methodological suggestions on how to more thoroughly explore, or just better control for, the role that autocorrelation can play (Section 2.1) and how to improve our statistical modeling – regression or otherwise – of alternation data (Sections 2.2 and 2.3) whereas Section 3 surveys a variety of predictors that alternation research is not incorporating enough. Section 4 will conclude.

2. Improvements 1: Better methods

2.1 More on autocorrelation

As discussed in the previous section, what happened the last time the speaker had to make a certain choice can already account for a large amount of variability in the choices that speakers make unconsciously by either blocking or facilitating the same choice next time around. In this section, I will discuss three additional aspects of (facilitatory) priming that much alternation research does not yet take into consideration.

2.1.1 *Beta persistence*

One particularly interesting kind of priming/persistence effect has been discussed in Szmrecsanyi (2005, 2006). He distinguished two kinds of priming effects: (i) α-persistence, the 'common' kind of structural priming effect, in which the use of a structure X facilitates the re-use of the same structure, and (ii) β-persistence, in which the use of a structure X facilitates the use of a similar/related structure Y. Examples of the former are passives priming passives, analytic comparatives priming analytic comparatives, *going-to* futures priming *going-to* futures, etc.; examples of the latter are how uses of *more* outside of analytic comparatives prime analytic comparatives, how uses of *go* as a motion verb prime *going-to* futures etc. This latter kind of priming, β-persistence, is therefore not purely structural priming, but cuts across lexical and structural levels. While that may make it harder to explain its strength, duration, and interactions with other factors, the fact that Szmrecsanyi found significant effects of α- and β-persistence on top of all other predictors in his case studies means that, for instance, studies of future tense choices need to involve not just whatever predictors we know affect future choices as well as what future tense was used the previous time around (α-persistence), but also where in the previous contexts, say, *go* was used as a motion verb and/or where *will* was used not in its future tense meaning (β-persistence).

2.1.2 *Cumulative priming*

Unfortunately, the situation is even more complex than what the last sentence of the previous section suggests. By now there has been quite a bit of work indicating that priming can be cumulative – within a conversation as contained in a corpus file or within an experiment. An early exploration of this was Scheepers (2003), who explored long-term priming within an experiment by splitting the data into an early and a late half, but did not find a significant difference between the two experimental halves. However, Jaeger and Snider (2008) study voice and *that*-relativizer omissions in corpus data and do find an effect of a variable they call cumulativity, viz. "the number of primes of each structure previously encountered or produced by the speaker […] (excluding the most recent prime)". Similarly, Gries and Wulff (2009) study *to* vs. *ing* complementation in the L2 English of L1 speakers of German and find a suggestive tendency for within-subject-accumulative priming in a sentence-completion experiment. Doğruöz and Gries (2012) find that speakers of Turkish become more accepting of unconventional syntactic expressions in experimental sessions involving no more than eight experimental stimuli; see also Francom (2009) for more discussion. This in turn means that not only does one need to include speakers' previous choice in one's account of their present choice (Section 2.1.1), it is in fact necessary to include speakers' previous choices by adding, for instance, a cumulative priming predictor that quantifies to what degree speakers have leaned towards a certain choice so far; note that this would not be redundant in a mixed-effects model because a speaker-specific intercept alone, for instance, would not adjust for the fact how speakers' preferences change over the course of an experiment.

2.1.3 *Surprisal*

The last two sections discussed what kinds of priming effects can affect alternation choices – this and the next section will briefly discuss two variables that moderate priming effects, i.e. strengthen or weaken them. The first of these is the notion of surprisal. Following Hale (2001: 4), who in turn refers back to work as early as Attneave (1959), surprisal can be defined as "the combined difficulty of disconfirming all disconfirmable structures at a given word" or, more mathematically, $\log_2 (1/ p\,(\text{word i} \mid \text{word } i\text{-}1, \ldots, \text{context}))$ or $-\log_2 p\,(\text{word i} \mid \text{word } i\text{-}1, \ldots, \text{context})$; thus, surprisal is a heuristic measure correlated with processing difficulty. With regard to priming, Jaeger and Snider (2008) study the voice alternation and add a predictor to their regression model which quantifies not how much the present *target* verb 'likes' actives or passives – that would just be the collostructional preference of a verb to the construction – but how much the *priming* verb 'likes' actives or passives. Intriguingly, they find that "[a]ctive-biased prime verbs appearing in the passive make the target more likely to be a passive than passive-biased prime verbs" (Jaeger and Snider 2008: 1063); in other words, if a prime verb is not used in the structure it usually prefers, i.e. is surprising in its use, then that constructional choice primes more strongly than if a prime verb had been used in the structure it usually prefers. Thus, a researcher would need to include not only each current choice and the preference of the lexical item whose future choice is at issue, but also the previous choices (for cumulativity) and the degree to which the previous choice was surprising.

2.1.4 *Similarity*

The final priming-related predictor that is not considered much in most alternation research yet is that of prime-target similarity. It has been recognized for a long time that strength of priming is affected by the so-called lexical-identity boost (Pickering and Branigan 1998, Gries 2005, Szmrecsanyi 2005): Priming from one construction in an (earlier) prime to next constructional choice in a (later) target is stronger when, for instance, prime and target share the same verb. That is, leaving aside all other predictors, a ditransitive with *give* is more likely to lead to another ditransitive when the second ditransitive will also involve *give* (as opposed to, say, *hand*). However, since then the role of similarity has been found to be more profound in how it operates on more levels than just verb identity. Two examples shall suffice.

First, similarity does not only have an effect when measured as lexical identity of prime and target as mentioned above – similarity also includes verb sense: Bernolet, Colleman, and Hartsuiker (2014) show that, in addition to the lexical identity boost, there is also an effect of verb-sense-identity boost that affects priming in the dative alternation such that priming is stronger when not just the verb, but also the verb sense is the same in both prime and target.

Second, there is an even more general effect of global prime-target similarity: Snider (2009) adopts a very global approach to similarity by comparing each prime to the corresponding target using the multi-feature Gower metric as a distance measure, which can compare the similarity of two objects based on many categorical and/or numeric features. He then tests the hypothesis that "two exemplars that are more similar in the sense that they share more features and have a lower [distance] between them, are more likely to prime" (p. 818) and indeed finds that "[w]hen the prime construction is PO [Prepositional Object, STG], the PO construction is 10.6 times more likely in the target for every one-unit decrease in [GlobalSim]" (p. 819).

2.1.5 *Interim summary*

In sum, just including autocorrelation and its associated effects properly requires a lot more annotation than is customarily done at this point because, in a sense, much work has been focusing on the context/situation of the *current* choice when much of what determines the current choice is not in the 'match column' of our concordance lines, it's in the 'preceding-context column' of our concordance lines and co-determined by *previous* choices: We need all predictors that traditional research has identified for all uses in a conversation/file/… so we can (i) enter them into a regression equation and (ii) compute the pairwise similarities of all prime-target pairs; we need to annotate previous contexts for β-persistence; we need the preferences of the relevant lexical items in the target but also in the prime (for surprisal); we need for every target a predictor that summarizes all previous choices in an index (for cumulativity), … the picture is much more complicated than nearly all existing work has been able to accommodate (and yes, that includes virtually all my own work on alternations, too). However, this also means that the subsequent statistical analysis will have to be complex and comprehensive to identify the relevant effects in an attempt to maximize (i) predictive accuracy but also (ii) explanatory power in the sense of being able to interpret the results, which will be the topic of the next section.

2.2 More on regression modeling

As mentioned above, the statistical method of choice for most contemporary alternation studies is some sort of regression model – by now most often a mixed-effects model with some adjustments for repeated measurements per speaker and/or item, but even 'regular' fixed-effects-only are still common (and often do not differ *that* much from their more advanced counterparts, which is to say that often they differ in effect size but rarely in effect direction). Again, this is a very welcome development in how such models can handle multiple predictors, their interactions, speaker/word-specific idiosyncrasies, and more all in a framework that is shared across many disciplines. Nevertheless, as the field has been developing this way, some recommendations from statistics or other disciplines have been heeded, but others have not, leaving the field in what sometimes looks like a strange blend of, on the one hand, advanced regression modeling practices and, on the other hand, sub-optimal practices and evaluation patterns that could benefit from some corrections/improvements, and I will briefly discuss some of these in this section.

2.2.1 *On model selection and model amalgamation*

One central question that nearly all regression-based studies face in one way or the other is that of which model's results to report and interpret. Much work, including my own, uses a model selection approach, in which the researcher begins with some initial model and then uses some multi-step strategy and diagnostics to arrive at 'the final model', whose significance tests, effect sizes, etc. are then reported (and, ideally visualized!). A lot of times, this model selection process involves (i) an initial model that is (relatively) maximal in both its fixed and random effects (e.g., it might contain all possible independent variables and all their 2- and 3-way interactions) and (ii) a stepwise process of backwards model selection, i.e. predictors are eliminated if they do not make enough of a contribution to the model (in terms of a significant *p*-value or, sometimes, *AIC* or *AIC* $_{c}$, Akaike's Information Criterion, see Harrell 2015, Section 9.8.1). While variations of this are possible – some studies would not just test predictors in the model for elimination, but also predictors not in the model for inclusion – this is the template that seems to underlie most studies.

While this approach has yielded very many insightful results, from a statistical perspective, it leaves some room for improvement especially when the model selection process is based on successive Likelihood Ratio tests or other *p*-values. First, researchers should be very aware of the fact that this approach is a somewhat awkward blend of exploratory analysis (the identification of a best model from the data by successively testing alternatives) and null hypothesis significance testing (NHST) (by using *p*-values, which quantify the probability of the data if H_0 was true). As has been known for a long time, this approach is problematic (see Harrell 2015: Section 4.3 for a very good summary) and the field would benefit from considering alternatives or, minimally, a heightened awareness of what this means for our studies.

One small improvement would be to rely less on significance tests and more on other criteria such as *AIC* or, even better, *AIC* $_{c}$. While using *AIC* $_{(c)}$ does not necessarily address all problems of model selection, it does avoid the pitfalls of the NHST paradigm, it affords the analyst more flexibility since he can compare not just nested, but also non-nested, models, and it allows the analyst to compute evidence ratios, a more useful way of comparing model quality than typically relatively meaningless *p*-values (see Burnham and Anderson 2002, Section 2.10).

A probably even better approach – to the extent it is computable for a certain data set (see below) – is offered by multimodel inferencing (see Burnham and Anderson 2002 for a general introduction and Kuperman and Bresnan 2012 for what might be the first application in linguistics). This method involves fitting many (motivated!) models to a data set at the same time and then, rather than selecting one 'best model', make inferences and predictions on the basis of a weighted combination of (the best of) all models, where *best* can be defined on the basis of the differences of all models' *AIC* $_{c}$-values from the best model's *AIC* $_{c}$-value. This process, called model averaging, does away with the assumption that there *is* one best model, it can be less affected by (moderate degrees of) collinearity, and suffers less from confirmation bias than the by now predominant model selection methods.

2.2.2 *On effects and comparisons*

Improving regression modeling does not end with the best choice of a model selection process or the most ideal model averaging/amalgamation because even the ways in which predictors are entered into the first or even only model can often be improved upon as well. One set of recommendation involves practitioners' treatment of numeric predictors in regression models. Most studies implicitly (appear to) assume that the effect of a numeric predictor can be characterized best by a straight line (maybe after a transformation using logs, square roots, inverse logits, …) or, if a straight line is considered sub-optimal, that non-linearities can be captured well by factorizing the numeric predictor into a categorical variable with multiple levels. It is probably fair to say that neither assumption is merited in most cases: true straight-line relationships are rare, and factorization of numeric predictors is often problematic; see Harrell (2015: Section 2.4). It would therefore be advantageous to (i) do away with factorization in most cases and (ii) explore curvature using regression splines, polynomials of numeric predictors, and/or use generalized additive (mixed) models (see Baayen et al., to appear a, b)– with these methods, growth/change curves, *U*-shaped developments, effects reaching asymptote etc. can be handled better than what many studies have worked with.

Another big improvement can be obtained by considering the random-effects structure in data more carefully. This is true in two ways: one is concerned with the 'size' of the random-effects structure and a highly influential paper (Barr et al. 2013) makes a variety of recommendations, which amount to trying to use a maximal random-effects structure, i.e. varying intercepts and slopes for all (non-control) predictors of interest; see Bates et al. (subm.) and Matuschek et al. (subm.) for responses. However, most discussions of random-effects structures in data focus on *crossed* random effects of the type encountered in carefully designed (psycholinguistic) experiments – there is much less discussion of the kind of *nested* random-effects structures frequently found in observational data. For instance, in statistical analyses of corpus data it would not be surprising to have to define varying intercepts for speakers, where the speakers are nested into files, with the files being nested into sub-registers, which are nested into registers, which are nested into modes (speaking vs. writing) and ideally one's analysis would take this multilevel structure into consideration. Similarly, newspapers in corpus data might be nested into dialectal/geographical varieties (Gries and Bernaisch 2016), annotators are nested into corpora (Gerard, Keller, and Palpanas 2010), etc. Thus, in order to do justice to the complexity of the data, more research needs to take these kinds of nested effects into consideration; see Gries (2015) for an intro for corpus linguistics and Judd, Westfall, and Kenny (2017) for a general and more technical introduction.

The other main way in which random effects need to be considered more carefully has to do with using the information they provide. Nearly always, researchers only utilize random effects by benefiting from how they make fixed-effects regression coefficients more precise – but random effects can also provide interesting information that may guide post-hoc data exploration. Miglio et al. (2013) use a mixed-effects model with varying intercepts to study experiencer choice (accusative vs. oblique) in data from the Corpus del Español. Interestingly, the varying intercepts as well as the predicted experiencer choices were significantly correlated with the authors' geographical regions (as operationalized with eight different areas). While this is just a simple example, it does show that it is possible for the speaker- or item-specific variation that is captured in the random-effects structure of a model to give rise to additional generalizations – disregarding them as a matter of habit does not utilize their patterning best.

Then, there is the question of how regression results for categorical predictors are computed and discussed. The default scenario these days is that scholars use the programming language R for their regression modeling, which in turn by default uses treatment contrasts where all levels of a categorical predictor are compared to a reference level, which is typically the alphabetically first level; what is then reported are the regression coefficients representing these contrasts as differences in the dependent variable resulting from different factor levels. There are several issues with that approach. First, sometimes authors do not state their reference level so regression coefficients are harder to interpret than necessary, but that's a very minor reporting problem only. Second and more importantly, (i) treatment contrasts are not orthogonal, i.e. not independent of each other and (ii) unless the point of the exercise is only to compare one group (e.g., a control group) to all others, they do not test (all) hypotheses or differences of interest. The former can lead to anticonservative significance tests because the regression coefficients are correlated and the latter misses out on the opportunity to test multiple meaningful *a priori* hypotheses at the same time. For instance, Hasselgård and Johansson (2011) study the frequency of *quite* in the English of native speakers and learners with German, Norwegian, French, and Spanish as their L1s. While they do not discuss the data with a regression model, this is a perfect example to exemplify my point. The kind of regression analysis that would probably be done with this data is to use a Poisson regression, make the L1 data the reference level, and then get regression coefficients for the following contrasts: $\text{L1}_{\text{English}}$ vs. $\text{L1}_{\text{German}}$, $\text{L1}_{\text{English}}$ vs. $\text{L1}_{\text{Norwegian}}$, $\text{L1}_{\text{English}}$ vs. $\text{L1}_{\text{French}}$, and $\text{L1}_{\text{English}}$ vs. $\text{L1}_{\text{Spanish}}$. However, not only are those contrasts not independent of each other, they also fail to test what Hasselgård and Johansson claim as one of their main findings, a "Germanic-Romance distinction." Thus, it would be much better to test the following four orthogonal contrasts:

(1) $\text{L1}_{\text{English}}$ vs. ($\text{L1}_{\text{German}}$ and $\text{L1}_{\text{Norwegian}}$ and $\text{L1}_{\text{French}}$ and $\text{L1}_{\text{Spanish}}$) i.e. native vs. non-native

(2) ($\text{L1}_{\text{German}}$ and $\text{L1}_{\text{Norwegian}}$) vs. ($\text{L1}_{\text{French}}$ and $\text{L1}_{\text{Spanish}}$) i.e. Germanic vs. Romance

(3) $\text{L1}_{\text{German}}$ vs. $\text{L1}_{\text{Norwegian}}$

(4) $\text{L1}_{\text{French}}$ vs. $\text{L1}_{\text{Spanish}}$

Not only are these contrasts uncorrelated, they are also more useful: They test whether the L1 data are different from all the L2 data, they test whether there is indeed a significant Germanic-Romance distinction (and it indeed turns out there is not), and they test whether *Germanic* and *Romance* are (statistically) homogeneous category levels. This approach is useful more often than one might think and can be applied usefully for many categorical predictors that have multiple levels and are often used in alternation studies, such as definiteness, animacy, person, degrees of literalness, etc. (A useful alternative for doing such comparisons *a posteriori* would be the framework of general linear hypothesis tests; see Bretz, Hothorn, and Westfall 2010.) In addition to testing multiple theoretically interesting *a priori* hypotheses, both approaches – *a priori* orthogonal contrasts and *a posteriori* general linear hypothesis tests – are useful for model selection: The vast majority of scholars adopting some model selection approach reduce the complexity of their models in accordance with Occam's razor in terms of independent variables and their interactions, but hardly anyone applies the same logic to the levels of the independent variables in the model, which is somewhat unexpected because both follow from Occam's rule: a model with fewer regression coefficients to be estimated is simpler than one with more regression coefficients regardless of whether the number of regression coefficients becomes smaller by dropping independent variables/interactions or factor levels! For instance, are all seven animacy levels one annotated supported by the data? Maybe a simpler model does (nearly) just as well in terms of classification/prediction but also avoids data sparsity and model convergence problems … All these are relatively minor tweaks with great potential to better meaningful hypotheses, avoid correlated estimates, and simplify models and their interpretation.

2.2.3 *On predicting and imputing*

The final few comments regarding regression modeling are concerned with how different regression models can be combined insightfully in alternation research in an approach referred to as MuPDAR (for Multifactorial Prediction and Deviation Analysis using Regressions). MuPDAR and derivatives has been fruitfully applied in particular in learner corpus research and variety research to illuminate how exactly linguistic choices of one or more target varieties (second/foreign language learners or speakers of a variety undergoing indigenization) differ from choices made of speakers of a reference/standard level (e.g., L1 speakers or speakers of a historical source variety; see Gries and Adelman (2014) or Gries and Deshors (2014) for the first applications.

MuPDAR borrows part of its logic from missing-data imputation: Often we have corpus data for what, say, L1 speakers have said in certain situations and corpus data for what L2 speakers have said in certain situations, but what we don't have and cannot cheaply/easily obtain is what L1 speakers would have said in the exact same situation that the L2 speakers were in. Yes, one could have each L2 speaker choice annotated by multiple L1 speakers for whether it is what they would have said, too, or whether they consider the L2 choice a mistake (i.e., perform error tagging/annotation), but not only would this add an enormous amount of work, it would also require dealing with (i) interrater reliability, (ii) intrarater reliability (do raters' ratings change as they see more input because of learning/habituation/satiation?), and (iii) the fact that one can ask raters only for a certain degree of fine resolution before the ratings become too fine-grained and, thus, potentially erratic. Thus, what MuPDAR does is it imputes the choices and/or judgments of the speakers of the references/standard; it does this with the following four-step procedure:

– one fits a first (regression, but see below) model on the reference/standard data to try and predict the choices of these speakers as well as possible; this model may not need to be interpreted so all that counts here is classification/prediction accuracy and robustness/lack of overfitting – model complexity, collinearity, etc. are less relevant here than if this was the only regression one fit on the data set as a whole;
– if this first model results in a good classification/prediction accuracy, then one applies this model to the target variety/varieties to impute, i.e. predict, what speakers of the reference/standard would have said;
– one can then compare for each choice of, say, a learner, whether it is identical to what a native speaker was predicted to say, and …
– … fit a second (regression, but see below) model on where the choices of the learners differ from the imputed native speaker choices and why (using minimally the same predictors as in the first model).

Crucially, most work using this approach has been using regression modeling – fixed-effects and mixed-effects modeling – but any other classification approach such as random forests etc. are also possible (see Deshors and Gries 2016); strictly speaking, one could even use a method that is very hard to interpret for the first model (e.g., support vector machines or neural networks) because the main purpose of that model is prediction – the results need not be particularly (easily) interpretable. Also, in particular for the first model, but actually for all kinds of regression modeling, cross-validation could be a useful addition to the method to make sure the model imputing the linguistic choices of the reference/standard speakers is robust and does not suffer from overfitting. This approach, therefore, allows one to get the most out of corpus data and illuminate even very subtle differences in how variables affect alternation choices in different speaker groups of interest.

2.3 Alternatives to regression modeling

This last section is a brief encouragement to readers to consider alternatives to the current standard of regression modeling, which has probably emerged as a standard because of how widely used regressions are in many fields, the high degree of sophistication to which they have evolved, and their relative ease of interpretability, but there are also arguments against them, or arguments in favor of other approaches. Three main considerations for using other methods have been advanced.

One is that regression modeling is not cognitively plausible, which is why methods such as naïve discriminative learning (Baayen 2011) and/or memory-based learning (Theijssen et al. 2013) and others might be more useful – if cognitive realism is what one desires *and* if one assumes that any of these methods *truly* reflects/does what the mind does, that is (a not uncontroversial assumption) …

Another one is that regression modeling sometimes does not actually yield the best classification/prediction accuracies so methods that score higher on those metrics may be more desirable such as support vector machines or more sophisticated kinds of neural networks, … even if those score lower on interpretability.

Finally, regression methods may simply not be feasible because, for instance, they make assumptions regarding the form of the data that are often not met, they can be affected much by collinearity and data sparsity (both in general or from combining predictors with Zipfian distributions), and they may sometimes just not be computable. For instance, Deshors and Gries (2016) gave up on regressions when data sparsity and collinearity ruled out 'normal' model selection and a multimodel inferencing script did not terminate within 200 hours on seven cores of an Intel i7 3.4 Ghz processor with Multithreaded BLAS libraries … In such cases and when 'cognitive realism' is not required, alternative methods include classification and regression trees (CART) or their above-mentioned bootstrapped cousin of random forests can be useful; also, as a side remark, scholars who reject frequentist statistics because of the many theoretical issues coming with the null hypothesis significance testing paradigm may prefer Bayesian modeling. Random forests are less likely to overfit, include a cross-validation aspect to, and usually return a higher accuracy than 'simple' classification and regression trees, but are harder to interpret so it is up to the analyst to prioritize and choose what is most appropriate for the task at hand.

In sum, the advent of ever increasing amounts of data and the recognition of the sometimes enormous degrees of complexity of data in alternation research require that linguists become more and more statistically and analytically savvy. While notions such as *big data*, *machine learning*, etc. have become a staple in computer science, statistics, and computational linguistics for a while, it is increasingly necessary even for descriptive and theoretical linguists to develop some familiarity with all of the above. This is especially the case given that newer alternation research has now also begun to widen the number and range of predictors that can or even should be considered, which is the topic of the next section.

3. Improvements 2: Underused predictors

Much research on syntactic alternations has shown that many alternations can be predicted fairly well on the basis of a small set of predictors:

- length-based predictors such as length in characters, syllables, morphemes, words, etc., which are all highly correlated with each other and, a bit less so, with some complexity metrics; in some cases, two lengths might be expressed in one predictor (as a difference or a ratio);
- givenness/activation from preceding discourse based on counts when, if at all, and how often a referent has been mentioned in the preceding discourse; while most approaches use lexical and/or referential identity, Gries (2003) showed that cover terms and part-whole relations may also have to be considered;
- NP type in general, but also definiteness and/or specificity;
- semantics of the main verb or the process denoted by the main verb or the relation of NPs (e.g., the semantic relation of possessors and possessums in the genitive alternation);
- animacy and/or concreteness and/or related predictors such as (degree of) idiomaticity or literalness;
- priming/persistence (see above); …

Obviously, many of these predictors are often highly correlated, e.g., prototypical and frequent recipients in the ditransitive will be short given definite pronouns referring to (concrete) human beings, with the whole VP denoting literal transfer of possession of the patient from the agent to the recipient. However, newer research has discovered a variety of other predictors that are much less studied, but are still relevant in at least some contexts and this section will very briefly discuss three kinds of them.

3.1 Information-theoretic predictors

Above, we have seen how the notion of surprisal is relevant for priming, but surprisal can also be at work in a different way. For example, Jaeger (2011) studies subject-extracted relative clauses (SRC), comparing reduced and full versions such as *The phaser used was set to 'stun'* and *The phaser that was used was set to 'stun'* respectively. He finds that using the full version is highly significantly correlated with the surprisal of seeing an SRC given the noun as well as the surprisal of seeing an SRC given the participle. In a similar application, Linzen and Jaeger (2015) find that the entropy (i.e. uncertainty) reduction of potential parse completions is correlated with reading times of sentences involving the DO/SC alternation between SC cases such as *Worf accepted Picard was right* and *Worf forgot Picard was right*: The complementation patterns *accept* takes have a lower entropy than the ones *forget* takes. Finally, Wulff, Gries, and Lester (to appear) find that *that*-complementation – the alternation of *I thought (that) the first officer likes the counselor* – is significantly correlated with the surprisal of the beginning of the complement clause (here, *the*) given the verb of the main clause (here, *thought*). In other words, an important class of predictors has to do with planning the upcoming constituents and how predictable each of the alternants is given the already processed material, but with the exceptions of a whole host of interesting studies by Jaeger, Levy, or Moscoso del Prado, this class of predictors has yet to have the desired impact on mainstream alternation research.

3.2 Phonological predictors

One particularly interesting class of predictors that are usually understudied – apart from persistence and its counterpart of *horror aequi* from above – are phonological in nature, and these are particularly interesting because, to the extent they can be shown to be correlated and maybe even causally related, they provide evidence for a view of language production in which supposedly late articulatory planning can have an impact on supposedly early morphosyntactic planning. In this section, I briefly point out two such predictors and how they have been used in previous work.

3.2.1 *Rhythmic alternation*

One interesting phonological predictor that few studies include is that of rhythmic alternation (Couper-Kuhlen 1986: 60), a principle that essentially states that sequences of two adjacent stressed syllables (stress clashes) as well as three or more adjacent unstressed syllables (stress lapses) are typically dispreferred. This principle has been shown to affect morphological variation (Schlüter 2003), but it is in fact also correlated with syntactic variation. For instance, in a study of particle placement – *John picked up the book* vs. *John picked the book up* – Gries (2007) finds in a series of monofactorial tests that

- with monosyllabic verbs and particles, stress clashes are significantly avoided;
- with disyllabic verbs of the structure stressed-unstressed and monosyllabic particles, V Prt NP is significantly preferred (given how it leads to a nice alternating pattern);
- sequences of verbs with final stress and particles with initial stress are significantly dispreferred (given how they lead to stress clashes); and more.

In a more useful multifactorial regression context, an operationalization would compute a score whose sign/amount represents how much 'more ideal' the chosen sequence of syllables is than the non-chosen alternative. While there is not yet much work out there that considers rhythmic alternation as a predictor, Schlüter (2015) as well as other papers in Vogel and Vijver (2015) and Wulff and Gries (2015) are a promising beginning.

3.2.2 *Segment alternation*

A similar predictor involves segment alternation, or ideal syllable structure (ISS), according to which sequences with CV alternations should be preferred over, say, sequences with many consonant clusters (Venneman 1988, Schlüter 2003). Wulff and Gries (2015) apply MuPDAR to prenominal adjective order – *lovely bright eyes* vs. *bright lovely eyes* – of L1 and L2 (German and Chinese) speakers of English and include ISS as a predictor (again as a score whose sign/amount reflects how much 'more ideal' the chosen sequence of words is over the alternative; for instance, *lovely bright eyes* should be preferred over *bright lovely eyes*). They find that not only are L1 speakers' adjective orderings compatible with ISS, L2 speakers' adjective orderings are, too, and Chinese learners' ordering choices are more in line with ISS than the German learners' ones, which may be due to the fact that German is much more tolerant of consonant clusters than Mandarin Chinese (see Li and Thompson 1981: 3).

3.3 Interim summary

While this section was by necessity brief, it is important to realize how the field is slowly going beyond the many predictors that have been used in many previous 20th century studies. I focused on information-theoretic and phonological predictors here because of how the former are currently a 'hot new thing' in much psycholinguistic work and because of how the latter are vastly understudied – this is of course not to say, however, that other factors such as lexical density, clause juncture, associations of specific lexical items or their senses to particular constructions etc. should not also find their way into empirical analyses more often.

4. Concluding remarks

The main conclusion of this paper is that morphosyntactic-alternation research has still some way to go to mature. While great strides have been made as the field has moved more and more towards multifactorial studies, much needs to be done: we need more awareness of

- how choices in an alternation context are not just a function of the current speech situation, but how what happened in the minutes before the current choice can also have a strong impact on what will happen next;
- how the currently most widely-used tool – (mixed-effects) regression modeling – is used and understood best, what pitfalls need to be avoided (from 'big' theoretical issues such as conflating exploratory and hypothesis-testing work in model selection to nitty-gritty details such as which contrasts to use and how to accommodate curved trends), and what extensions of, and alternatives to, regression modeling to consider and why;
- the wide range of predictors that are potentially relevant and are included to get the best understanding of the alternations we study.

The bad news, as so often, is that this requires an ever increasing or nearly overwhelming degree of sophistication and knowledge not only in linguistics but also in matters of data analysis and statistical evaluation. The good news, though, is that this degree of complexity reflects how far the field has come in appreciating the complexities of our multidimensional data, and that making these next steps will hopefully take alternation research to the next level.

References

Attneave, Fred. 1959. *Applications of Information Theory to Psychology: A Summary of Basic Concepts, Methods and Results*. Holt, Rinehart and Winston.

Baayen, R. Harald. 2011. "Corpus Linguistics and Naïve Discriminative Learning." *Brazilian Journal of Applied Linguistics* 11(2): 295–218.

Baayen, R. Harald, Jacolien von Rij, Cecile de, Cat C., and Simon N. Wood. to appear a. Autocorrelated Errors in Experimental Data in the Language Sciences: Some Solutions Offered by Generalized Additive Mixed Models. In *Mixed effects regression models in Linguistics*, ed by Dirk Speelman, Kris Heylen, and D. Geeraerts. Berlin and Springer.

Baayen, R. Harald, Shravan Vasishth, Douglas M. Bates, and Reinhold Kliegl. to appear b. "The Cave of Shadows. Addressing the Human Factor with Generalized Additive Mixed Models." *Journal of Memory and Language*.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. "Random Effects Structure for Confirmatory Hypothesis Testing: Keep it Maximal." *Journal of Memory and Language* 68(3): 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, Douglas M. Reinhold Kliegl, Shravan Vasishth, and R. Harald Baayen. submitted. "*Parsimonious Mixed Models*."

Bernolet, Sarah, Timothy Colleman, and Robert Hartsuiker. 2014. "The 'Sense Boost' to Dative Priming: Evidence for Sense-Specific Verb-Structure Links." *Journal of Memory and Language* 76(1): 113–126. https://doi.org/10.1016/j.jml.2014.06.006

Bretz, Frank, Torsten Hothorn, and Peter Westfall. 2010. *Multiple Comparisons Using R*. Boca Raton, FL, London, and New York: Chapman and Hall / CRC. https://doi.org/10.1201/9781420010909

Burnham, Kenneth P. and David R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed. London and New York: Springer.

Couper-Kuhlen, Elizabeth. 1986. *An Introduction to English Prosody*. Tübingen: Edward Arnold and Niemeyer.

Deshors, Sandra C. and Stefan Th. Gries. 2016. "Profiling Verb Complementation Constructions across New Englishes: A Two-Step Random Forests Analysis to *ing* vs. *to* Complements." *International Journal of Corpus Linguistics* 21(2): 192–218.

Francom, Jerid. 2009. Experimental syntax: Exploring the Effect of Repeated Exposure to Anomalous Syntactic Structure: Evidence from Rating and Reading tasks. Ph.D. dissertation, University of Arizona.

Gerard, Jeffrey, Frank Keller, and Themis Palpanas. 2010. "Corpus Evidence for Age Effects on Priming in Child Language." In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, ed. by Stellan Ohlsson and Richard Catrambone, 1559–1564.'

Gries, Stefan Th.. 2003. *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. London and New York: Continuum Press.

Gries, Stefan Th.. 2007. "New Perspectives on Old Alternations." In *Papers from the 39th Regional Meeting of the Chicago Linguistics Society: Vol. II. The Panels*, ed. by Jonathan E. Cihlar, Amy L. Franklin, and David W. Kaiser, 274–292. Chicago, IL: Chicago Linguistics Society.

Gries, Stefan Th.. 2011. "Commentary." In Kathryn Allan and Justyna Robinson (eds.), *Current Methods in Historical Semantics*, 184–195. Berlin and New York: Mouton de Gruyter. https://doi.org/10.1515/9783110252903.184

Gries, Stefan Th.. 2015. "The Most Underused Statistical Method in Corpus Linguistics: Multi-Level (and Mixed-Effects) models." *Corpora* 10(1): 95–125. https://doi.org/10.3366/cor.2015.0068

Gries, Stefan Th.. 2016. "Frequencies of (Co-)Occurrence vs. Variationist Corpus Approaches towards Alternations: Variability due to Random Effects and Autocorrelation." In *Triangulating Methodological Approaches in Corpus Linguistic Research*, ed. by Paul Baker and Jesse Egbert, 108–123. New York: Routledge, Taylor and Francis.

Gries, Stefan Th.. and Allison S. Adelman. 2014. "Subject Realization in Japanese Conversation by Native and Non-Native speakers: Exemplifying a New Paradigm for Learner Corpus Research." In *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, ed. by Jesús Romero-Trillo, 35–54. Cham: Springer.

Gries, Stefan Th.. and Tobias J. Bernaisch. 2016. "Exploring Epicenters Empirically: Focus on South Asian Englishes." *English World-Wide* 37(1): 1–25.

Gries, Stefan Th.. and Sandra C. Deshors. 2014. "Using Regressions to Explore Deviations between Corpus Data and a Standard/Target: Two Suggestions." *Corpora* 9(1): 109–136. https://doi.org/10.3366/cor.2014.0053

Gries, Stefan Th.. and Anatol Stefanowitsch. 2004. "Extending Collostructional Analysis: A Corpus-Based Perspective on 'Alternations'." *International Journal of Corpus Linguistics* 9(1): 97–129. https://doi.org/10.1075/ijcl.9.1.06gri

Gries, Stefan Th.. and Stefanie Wulff. 2009. "Psycholinguistic and Corpus Linguistic Evidence for L2 Constructions." *Annual Review of Cognitive Linguistics* 7: 163–186. https://doi.org/10.1075/arcl.7.07gri

Hale, John. 2001. "A Probabilistic Earley Parser as a Psycholinguistic Model." *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*. https://doi.org/10.3115/1073336.1073357

Harrell, Frank E. Jr. 2015. *Regression Modeling Strategies. […]*. 2nd ed. London and New York: Springer. https://doi.org/10.1007/978-3-319-19425-7

Jaeger, T. Florian. 2011. "Corpus-Based Research on Language Production: Information Density and Reducible Subject Relatives." In *Language from a Cognitive Perspective: Grammar, Usage, and Processing*, ed. by Emily M. Bender and Jennifer Arnold, 161–197. Stanford: CSLI.

Jaeger, T. Florian and Neal Snider. 2008. "Implicit Learning and Syntactic Persistence: Surprisal and Cumulativity." In *Proceedings of the Cognitive Science Society Conference*, ed. by Bradley C. Love, Kenneth McRae, K., Vladimir M. Sloutsky, 1061–1066. Washington, DC.

Judd, Charles M., Jacob Westfall, and David A. Kenny. 2017. "Experiments with More than One Random Factor: Designs, Analytic Models, and Statistical Power." *Annual Review of Psychology* 68(1).

Kuperman, Victor and Joan Bresnan. 2012. "The Effects of Construction Probability on Word Durations during Spontaneous Incremental Sentence Production." *Journal of Memory and Language* 66(4): 588–611. https://doi.org/10.1016/j.jml.2012.04.003

Li, Charles N., and Sandra A. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, Los Angeles: University of California Press.

Linzen, Tal and T. Florian Jaeger. 2015. "Uncertainty and Expectation in Sentence Processing: Evidence From Subcategorization Distributions." *Cognitive Science* 40(6): 1382–1411.

Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, R. Harald Baayen, and Douglas M. Bates. subm. "Balancing Type I Error and Power in Linear Mixed Models."

Miglio, Viola G., Stefan Th. Gries, Michael J. Harris, Eva M. Wheeler, and Raquel Santana-Paixão. 2013. "Spanish *lo*(*s*)-*le*(*s*) Clitic Alternations in Psych Berbs: A Multifactorial Corpus-Based Analysis." In *Selected Proceedings of the 15th Hispanic Linguistics Symposium*, ed. by Jennifer Cabrelli Amaro, Gillian Lord, and Ana de Prada Pérez, and Jessi E. Aaron, 268–278. Somerville, MA. Cascadilla Press.

Pickering, Martin J. and Holly P. Branigan. 1998. "The Representation of Verbs: Evidence from Syntactic Priming in Language Production." *Journal of Memory and Language* 39(4): 633–651. https://doi.org/10.1006/jmla.1998.2592

Scheepers, Christoph. 2003. "Syntactic Priming of Relative Clause Attachments: Persistence of Structural Configuration in Sentence Production." *Cognition* 89(3): 179–205. https://doi.org/10.1016/S0010-0277(03)00119-7

Schlüter, Julia. 2003. "Phonological Determinants of Grammatical Variation in English: Chomsky's Worst Possible Case." In *Determinants of Grammatical Variation in English*, ed. by Günter Rohdenburg and Britta Mondorf, 69–118. Berlin, New York: Mouton de Gruyter. https://doi.org/10.1515/9783110900019.69

Schlüter, Julia. 2015. "Rhythmic Influence on Grammar: Scope and Limitations." In *Rhythm in Phonetics, Grammar and Cognition*, ed. by Ralf Vogel and Ruben Vijver, 179–206. Berlin: De Gruyter Mouton.

Snider, Neal. 2009. "Similarity and Structural Priming." In *Proceedings of the 31th Annual Conference of the Cognitive Science*, ed. by Niels A. Taatgen and Hedderik van Rijn, 815–820.

Stefanowitsch, Anatol and Stefan Th. Gries. 2003. "Collostructions: Investigating the Interaction between Words and Constructions." *International Journal of Corpus Linguistics* 8(2): 209–243. https://doi.org/10.1075/ijcl.8.2.03ste

Szmrecsanyi, Benedikt. 2005. "Language Users as Creatures of Habit: A Corpus-Linguistic Analysis of Persistence in Spoken English." *Corpus Linguistics and Linguistic Theory* 1(1): 113–150. https://doi.org/10.1515/cllt.2005.1.1.113

Szmrecsanyi, Benedikt. 2006. *Morphosyntactic Persistence in Spoken English. A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis*. Berlin and New York: Mouton de Gruyter. https://doi.org/10.1515/9783110197808

Theijssen, Daphne, Louis ten Bosch, Lou Boves, Bert Cranen and Hans van Halteren. 2013. "Choosing Alternatives: Using Bayesian Networks and Memory-Based Learning to Study the Dative Alternation". *Corpus Linguistics and Linguistic Theory* 9(2): 227–262. https://doi.org/10.1515/cllt-2013-0007

Vogel, Ralf and Ruben Vijver (eds.). 2015. *Rhythm in Phonetics, Grammar and Cognition*. Berlin, New York: De Gruyter Mouton.

Vennemann, Theo. 1988. *Preference Laws for Syllable Structure and the Explanation of Sound Change. With Special Reference to German, Germanic, Italian and Latin*. Berlin, New York: Mouton de Gruyter.

Wulff, Stefanie and Stefan Th. Gries. 2015. "Prenominal Adjective Order Preferences in Chinese and German L2 English: A Multifactorial Corpus Study." *Linguistic Approaches to Bilingualism* 5(1): 122–150. https://doi.org/10.1075/lab.5.1.05wul

Wulff, Stefanie, Stefan Th. Gries, and Nicholas Lester. To appear. "Optional *that* in Complementation by German and Spanish Learners: Where and How German and Spanish Learners Differ from Native Speakers." In *What Does Applied Cognitive Linguistics Look Like? Answers from the L2 Classroom and SLA Studies*, ed. by Andrea Tyler and Carol Moder. Berlin, Boston: De Gruyter Mouton.

Author's address

Stefan Th. Gries
Department of Linguistics
University of California, Santa Barbara
Santa Barbara, CA 93106-3100
United States of America

stgries@linguistics.ucsb.edu