

Some Implications of Competent Gradience
Nigel Duffield (University of Nijmegen & University of Sheffield)

0. Abstract

This paper is concerned with the phenomenon of gradience — patterns of systematic *variability* — in acceptability judgments, and with the nature of the underlying abilities that give rise to these gradient patterns in performance. It is also concerned with a particular kind of *variation*, namely, with differences between native-speakers and second language learners with respect to the weighting of implicit factors determining these gradient judgments. By drawing attention to the relationship between this type of *intra-speaker variability* and *inter-speaker variation* in *performance* with respect to theories of *grammatical competence*, it is hoped to make some contribution to a theory of syntactic variation.

1. Introduction: Competent Gradience and the Competence Paradox

In Duffield (2003), I set out to address how best to treat non-discrete, or intermediate, acceptability judgments, a theoretical and methodological issue that is largely ignored in current generative research. In that paper, I focussed attention on the phenomenon of COMPETENT GRADIENCE. This term was used to refer simultaneously to gradient patterns in acceptability judgments — a phenomenon of language performance — as well as to the implicit capacities that enable language users to make these reliably gradient judgments: that is, to mechanisms of linguistic competence.

As was observed, though perhaps insufficiently stressed, in that paper, the phenomenon of competent gradience in performance is independent of claims about the discreteness (or otherwise) of the underlying mechanisms. That is to say, it is essential to distinguish claims about the gradient nature of *acceptability judgments* from claims about the internal capacities —

including putative *grammatical constraints* — that give rise to these judgments.¹

However, the main contention of the previous article was that the phenomenon of competent gradience merits greater theoretical attention, since it can serve to distinguish native-speakers from second language learners in experiments investigating ultimate attainment, and hence may be considered — implicitly or explicitly — a benchmark of successful acquisition.

This is brought out especially clearly in cases in which second language learners ‘outperform’ native-speakers relative to some theoretically-determined set of judgments: instances of what I termed the COMPETENCE PARADOX. An illustration of this is provided by the sentences in (1) below, cited in Kluender (1992), and originally due to Chung and McCloskey (1983).

- (1) a. This is *a* paper that you really need to find someone that you can intimidate with. >
b. *Which paper* do you really need to find someone that you can intimidate with? >
c. *How many papers* do you really need to find someone that you can intimidate with? >
d. *What* do you really need to find someone that you can intimidate with?

All of the sentences in (1) are equally *ungrammatical* in the strict sense of violating parallel syntactic constraints — see fn. 1 — yet, intuitively, their

¹ Throughout this paper, I try to use the term ‘grammatical’ in the strict sense of ‘licensed — or generated — by the grammar’, where grammar refers simultaneously to a formal linguistic description in terms of rules and constraints, as well as to the putatively autonomous internalized representation of these rules and constraints (assuming the latter exists). I use the term ‘acceptable’ to designate utterances that language users would classify — either intuitively or explicitly — as ‘part of their language.’

status ranges from (almost fully) *acceptable* to strongly unacceptable. To be a competent 'knower of English', it is not sufficient to know that all these sentences are less than perfectly acceptable; one should also know that they are rather precisely ordered, each example being less acceptable than the one that immediately precedes it: this cline indicated by the '>' symbol. (In this case, the degrees of acceptability are a function of a semantic factor that Kluender 1992 terms 'referential specificity', rather than of syntactic structure *per se*.)

This is an instance of competent gradience: for whatever underlying reasons, competent knowers of the language judge these sentences neither as 'good' nor 'bad', but rather as 'increasingly less acceptable'; they also converge on the particular ordering involved.

The gradient acceptability judgments in (1) raise two larger questions about the empirical basis of so-called 'grammaticality judgments'. The first is whether it is ever possible to tease out any purely grammatical component of an acceptability judgment, given the non-structural factors that are always at play in determining acceptability. A second, more radical, question is whether what are claimed to be syntactic effects are always entirely determined by the interaction of non-structural factors; that is, whether (syntactic) grammaticality is merely an epiphenomenon of successful performance, as recent connectionist work appears to claim (see, for example, Allen and Seidenberg (1999), Seidenberg and MacDonald (1999)).²

² In the case of the sentences in (1), for example, the issue is whether something like the Subadjacency Principle — formulated to explain the putative *ungrammaticality* of these sentences — has any explanatory role in accounting for the cline in acceptability: if *not* — if all of the effect can be explained by non-structural factors — then it is reasonable to consider eliminating the constraint itself from the underlying competence grammar. The issue is

The Competence Paradox arises when we consider a hypothetical advanced L2 learner's response to such sentences. Given a categorical view of grammatical competence, an L2 learner who categorically rejected all of these sentences as unacceptable should in fact be more correct relative to the hypothesized grammar (*I-language*) than the English native speaker who is inclined to accept many or all of these sentences most or all of the time. Cases of L2 learners outperforming native-speakers in judgment tasks of this kind are not uncommon.

Intuitively, though, a learner who judges as *unacceptable* what native-speakers judge as *less* acceptable has failed to achieve native-speaker competence, if we define competence as whatever implicit linguistic knowledge determines successful language use.

Thus, by one measure the L2 learner whose judgments fail to match those of the native-speaker will be judged *more* competent than the native speaker, where competence is defined in terms of convergence between a pattern of overt behaviour and the hypothesised underlying categorical knowledge. Another measure, as well as common sense, dictates that the same L2 learner exhibiting the same response pattern should be judged *less* competent than the native-speaker — where competence is defined in terms of convergence of L2 learners on native-speakers' judgments. Although most generative second language researchers claim that they are probing the first type of competence, almost all work that I am aware of determines successful acquisition in terms of the second measure.³

whether the same applies to *all* such autonomous syntactic constructs. For some relevant discussion, see Culicover (1998).

³ It is reasonable to think that the Competence Paradox arises also with respect to intra-language variation. Consider, for example, a group of native-speakers who 'know' several interacting varieties of a given language, including the prestige variety in a local context, and a group of L2 learners taught only the high prestige standard. It is entirely possible that the

In Duffield (2003), I attempt to resolve this paradox by proposing a ‘dual competence’ model that incorporates two kinds of implicit competence: an UNDERLYING COMPETENCE, which *is* categorical — and putatively partially innate — and a more probabilistic SURFACE COMPETENCE, which encodes lexical and constructional contingencies (including lexical frequency, contextual appropriateness of a given construction, Kluender’s ‘referential specificity’, *etc.*).

Competent gradience, I claimed, arises through the interaction between these two sources of implicit knowledge, as well as through the interaction of this dual competence with other modules of language comprehension and production. The model was shown to explain a number of systematic mismatches in different language tasks among three ‘populations of judgments’: the offline introspective judgments of theoretical linguists; online and offline judgments of native-speakers; and online and offline judgments of second language learners.⁴

latter group will come to have ‘sharper’ intuitions about the grammar of the standard variety than the native-speakers.

Likewise, it is conceivable that the ‘dual competence’ approach could be adapted to model dialect and register variation, with common grammatical properties being assigned to underlying competence, and varying properties assigned to surface competence. While such a translation is undoubtedly much too simplistic, a more sophisticated adaption might have theoretical merit.

⁴ The terms ‘offline’ and ‘online’ are familiar in the psycholinguistics literature, but may be less so in the present context. Briefly, an offline task is one in which the experimental subject is under no time pressure to arrive at an acceptability judgment, and where the time-course of the decision process is irrelevant: traditional ‘grammaticality judgment tasks’ exemplify this type of task, whether conducted by an introspective armchair linguist consulting his/her own intuitions, or by a SLA researcher administering questionnaires. By contrast, in an online task subjects are typically under time pressure, otherwise response latency measures — often called reaction times — would be meaningless. More importantly, the time-course of the

Rather than recapitulating arguments presented previously, here I wish to consider more closely some of the ways in which intermediate acceptability judgments may arise, and to examine the implications of competent gradience for models of grammatical competence.

In order to provide empirical grounding for this discussion, I will focus on a specific set of experiments investigating a gradient phenomenon — namely, the Parallelism Effect in English VP-Ellipsis (VPE) — where the results from native-speakers contrast with those of second language learners in an illuminating fashion.

What is immediately at stake here is understanding how such gradient effects arise, how they are measured, and appreciating the significance of the differences that emerge between native-speakers and second language learners when their behaviour is carefully examined.

The rest of the paper is organised as follows. In the following section, I describe the Parallelism Effect and outline the main results of our previous experiments (Duffield & Matsuo 2003a, 2003b).⁵ In the final section, I discuss a possible interpretation of these results and consider their implications for current models of grammatical competence.

judgment process is crucial: psycholinguists are interested not only in *what types* of implicit information determine a given response, but also in *when* these different types of information make their contribution. Strictly speaking, the online experiments reported here are only ‘pseudo-online’, since the (only) probe point is at the *end* of each stimulus pair. This contrasts for example, with tasks such as self-paced reading, or eye-tracking, where measures are taken at different points in stimulus presentation.

⁵ This section, up to subsection 2.3.2, is a slightly modified version of part of Duffield and Matsuo (2003b). The subsections that follow in this section summarise part of Duffield and Matsuo (2003a).

2. On The Gradient Nature of the Parallelism Effect

The contrasting pairs of examples in (2) and (3) below illustrate the Parallelism Effect in English VPE constructions. According to the theoretical literature, a necessary condition for (grammatical) VPE is strict syntactic parallelism between the elided verb-phrase, indicated by the delta symbol, and its linguistic antecedent, indicated by italics:

- (2) a. Someone had to *put out the garbage*, since I hadn't Δ.
b. ??The garbage *had to be put out*, since I hadn't Δ.
- (3) a. It annoyed Mary when people *mentioned her sister's name*. Bill did Δ, quite often.
b. The *mention of her sister's name* always annoyed Sally. ??Tom did Δ, out of spite.

Since Sag (1976), there have been at least three conventional assumptions concerning this effect.⁶ The first assumption is that the Parallelism Effect results from a *categorical constraint*. That is to say, ellipsis clauses immediately following non-parallel antecedents are strictly ungrammatical, rather than being degraded for extra-grammatical reasons.^{7, 8}

⁶ See also Hankamer and Sag (1976), Sag and Hankamer (1984).

⁷ Notice that elided sentences have no inherent grammaticality status *per se*: in contrast to many other types of sentence, which can be judged in isolation, the relative acceptability of an ellipsis clause depends entirely on the form of the antecedent clause. This is not to suggest the advisability (or otherwise) of the standard practice of judging sentences in isolation, or to make any claims about the possibility of 'true null contexts'. The point here is much more basic: normal sentences *can* be judged in isolation, ellipsis clauses *cannot*.

⁸ Note also that I am using the term 'extra-grammatical' as complementary to the strict sense of 'grammatical' assumed throughout: extra-grammatical factors are still *internal*, cognitive factors, such as those that determine discourse cohesion. Here, I have nothing to say about external, non-cognitive factors determining the form of utterances.

Consistent with this assumption, sentences involving violations of the parallelism constraint are usually designated with an asterisk in the formal theoretical literature, to indicate their ungrammaticality. The use of an asterisk strongly implies that such sentences are also unacceptable, since ungrammaticality typically correlates with unacceptability, *absente* any independent processing or pragmatic difficulties. (This close correlation is for the natural — if unfortunately circular — reason that formal theories of grammaticality are themselves largely based on acceptability judgments; hence the misleading term 'grammaticality judgment').

The *empirical* question here is whether the effect induced by parallelism violations patterns with the effects of violations of other *syntactic* constraints, or whether it is more like those induced by extra-grammatical violations. A closely related, *theoretical* question is whether such effects are best interpreted in terms of an underlying autonomous grammatical constraint, or in terms of other, non-structural factors.

These questions are often conflated in formal analysis, but the first is a matter of fact, the second a matter of interpretation. In what follows, I will attempt to keep the two apart.

The second conventional assumption regarding the Parallelism Effect, related to the first, is that the effect is *uniquely* due to syntactic factors; that is to say, the underlying constraint alone explains the Parallelism Effect.

Finally, again related to the first assumption, it is assumed that the syntactic constraint — and hence the effect — applies *asymmetrically*, affecting VPE constructions while sparing very similar VP-Anaphora (VPA) constructions, such as those in (4) and (5):

- (4) a. Someone had to put out the garbage, since I hadn't done it.
b. ??The garbage had to be put out, since I hadn't done it.

- (5) a. It annoyed Mary when people mentioned her sister's name. Bill did it, quite often.
b. ?The mention of her sister's name always annoyed Sally. Tom did it, out of spite.

Our research, reported in Duffield & Matsuo (2003a, 2003b) challenges all three assumptions; see also Hardt (1993). In a set of psycholinguistic experiments, we examined non-structural factors that were hypothesised to significantly influence the Parallelism Effect, independently of any autonomous syntactic constraint.⁹ Our experimental results demonstrate systematic involvement of other factors — especially finiteness and construction type — in the acceptability of various kinds of parallelism violation, and suggest a more interactional interpretation of the underlying capacities determining this performance effect than is usually assumed.

To appreciate the significance of these results, it is useful to consider the methodology and methods of data analysis in some detail.

2.1. Teasing Apart Formal Factors

In our experiments, we manipulated various formal properties of the antecedent clause and of the ellipsis clause in ellipsis contexts.

In the antecedent clause, we manipulated CONSTRUCTION TYPE.¹⁰ Following earlier work by Tanenhaus and Carlson (1990), we contrasted two types of

⁹ I gratefully acknowledge the help of research assistants at the Max Planck Institute for Psycholinguistics, Nijmegen (MPI) and at McGill University in the completion of this project, especially Christopher Miller, Femke Uitdewilligen (MPI) and Ingrid Leung (McGill). Our research (Nigel Duffield & Ayumi Matsuo) was partially supported by a grant from Social Sciences and Humanities Research Council of Canada (410-98-0176, Lydia White P.I.), and internally, by MPI.

¹⁰ The experiment manipulated a second antecedent clause property, which we termed RECOVERABILITY: we were interested, for example, in whether the presence of a by-phrase in

non-parallel antecedent: passive, as in (2b), *vs.* nominal, as in (3b), above. Our prediction, based on naïve intuitions, was that the parallelism effect would be significantly weaker for passive than for nominal antecedents. Notice that standard generative accounts predict no difference in degrees of acceptability as a function of construction type. This is because, by hypothesis, construction-*specific* rules have no theoretical or psychological status: see, for discussion, Chomsky (1981), Chomsky (1986). In contemporary generative theory, for example, there is no ‘passive construction’: rather, passive sentences result from the interaction of autonomous general principles, such as Theta Theory and Case Theory, which themselves make no reference to specific constructions. Hence — logically, and in this experiment — construction-type and (autonomous) syntactic parallelism are independent variables.

In the ellipsis clause, we manipulated two further properties. The first was ANAPHOR TYPE, again replicating Tanenhaus & Carlson's experiments. Keeping the antecedent clause constant, we contrasted pairs of sentences with VPE completions, as in (2) and (3), with those with VPA completions, as in (4) and (5). Previous work suggested that the parallelism effect does not spare VPA completely: contrary to theoretical assumptions, there is reasonably strong intuitional and experimental evidence that non-parallel antecedents also degrade VPA completions (albeit less strongly than they do VPE completions).

It is worth pointing out why, from a theoretical perspective, violations of parallelism should *not* affect VPA. The classic account of the VPE *vs.* VPA

passive antecedents facilitated the interpretation of the ellipsis clause, this reducing the parallelism effect. Since the effects of recoverability were rather small relative to those of the other factors, it is omitted from the present discussion. See Duffield & Matsuo (2003a, 2003b) for further details.

dichotomy is due to Hankamer & Sag (1976), henceforth HS. HS propose a formal distinction between two types of anaphoric expression, which they term SURFACE and DEEP ANAPHORA, respectively. Surface anaphora refers to constructions, including VPE, whose grammaticality — which HS equate with acceptability — is claimed to be determined by the constraints on the syntactic form of the antecedent clause. In HS, the relevant syntactic representation is SURFACE STRUCTURE (hence the label).¹¹ Surface anaphora is contrasted with cases of deep anaphora, such as VPA, where the acceptability and intended interpretation of anaphoric elements are said to be derived directly from the discourse model, rather than *via* any syntactic representation of the antecedent. A consequence of this claim is that the syntactic form of the antecedent clause is irrelevant to the acceptability of deep anaphora constructions.¹²

The second ellipsis clause property we investigated was FINITENESS: our experiments systematically contrasted *finite* ellipsis clauses, headed by a form of the present perfect auxiliary have, with *non-finite* ellipsis clauses, headed by to. This contrast is illustrated in (6) below.

- (6) a. When we got back, our driveway had *been cleared of snow* — ??A neighbour told us that Tom had Δ .

¹¹ For a number of reasons, the authors later modify this claim. In Sag & Hankamer (1984), ellipsis constructions are claimed to refer to ‘propositional representations of the antecedent clause’, rather than surface structure; such representations are approximately equivalent to LF representations in other frameworks. The essential claim, which remains, is that ellipsis constructions are obligatorily sensitive to the *syntactic* properties of the antecedent clause; the structural parallelism constraint is one effect of this.

¹² In their 1984 paper, Sag & Hankamer also revise their characterization of deep anaphora constructions, recasting deep anaphora as ‘model-interpretive anaphor[a]’. However, the essential distinction remains, as do the consequent claims for the (non-)availability of parallelism effects.

- b. Our driveway needed to *be cleared of snow* — ?But no-one wanted to Δ .

In this case, informal intuitions — as well as prior results (Duffield and Matsuo (2001)) — indicated that non-finite violations of syntactic parallelism should be significantly more acceptable than finite violations; once again, the standard theory neither predicts nor readily accommodates such a contrast.

The factors manipulated in our experiments are summarized in Table 1 below. The number in the right-hand column represents the number of test pairs per cell in the design.

Syntactic Form/Construction Type	Ellipsis Clause	
	Finiteness	Anaphor Type
Parallel/Active-Verbal	Finite	VPE 16
		VPA 16
	Non-finite	VPE 12
		VPA 12
Non-Parallel/Passive	Finite	VPE 16
		VPA 16
	Non-finite	VPE 12
		VPA 12
Non-Parallel/Nominal	Finite	VPE 16
		VPA 16
	Non-finite	VPE 12
		VPA 12

Table 1. Summary of Formal Factors Investigated

2.2. Experiment

2.2.1. Subjects

We tested 31 native-speakers of North American English, and 20 Dutch-speaking (advanced L2 English) learners on the same materials.

We chose Dutch learners for several reasons. Most importantly, Dutch grammar does not license VPE constructions: thus, VPE is ungrammatical — and unacceptable — in Dutch, irrespective of the parallelism

of the antecedent clause. Consequently, if advanced Dutch learners of English show gradient sensitivity to violations of parallelism in English VPE, then the ability underlying this must be acquired in the course of second language development: there is no possibility of appealing to direct ‘L1 transfer’.

Furthermore, Dutch grammar permits VPA, which exhibits mild parallelism effects. This raises the question as to whether Dutch learners can acquire *differential sensitivity* to English parallelism effects, systematically preferring non-parallel VPA over non-parallel VPE — a categorical distinction — while at the same time treating both as ‘less than perfectly acceptable’ — a gradient, or intermediate, effect.

A second reason for choosing Dutch learners was more pragmatic: by comparison with many other learner groups, it is easy to find a relatively large, homogeneous group of speakers with a high enough level of proficiency to show sensitivity to this type of effect.

Subjects were divided into four groups: for each sentence-pair, each group received one of the four possible alternants: parallel-VPE, parallel-VPA, non-parallel-VPE, non-parallel VPA; see Table 1 above. There were a total of 128 test items per subject, plus 56 distracter/control items. For each run of the online experiments, all items were randomized within each block; for the offline experiment, the items were randomized separately in each version.

2.2.2. Methodology

Our experiment was a modification of the Sentence Completion Judgment Paradigm initially presented in Tanenhaus and Carlson (1990); see also Mauner, Tanenhaus and Carlson (1995). In this task, subjects are asked to judge, for pairs of sentences such as those in (2)-(6), whether the second (elided) sentence forms a ‘sensible completion’ to the first.

The experiment involved two separate tasks: an *online* timed anomaly task — replicating Tanenhaus & Carlson — and an *offline* Grammaticality Judgment task (using the same stimulus materials).

In the online task, the sentence-pairs were visually presented on a computer screen. For each stimulus-pair, subjects read the first sentence — containing the (parallel/non-parallel) antecedent clause, in the case of test sentences — then pressed a button on the keyboard when they were ready. The first sentence disappeared, and was replaced by the second sentence — containing the corresponding VPE or VPA clause). Subjects then had to accept or reject this second sentence as a ‘sensible completion’ as quickly as possible (by pressing a “yes” or a “no” button). In the offline version of the experiment, the two sentences were presented simultaneously in a written questionnaire.

The online task yielded two dependent measures: the proportion of sentences accepted in each condition (the *judgment data*); and the time taken to accept or reject the sentences of each condition (the *latency data*). The offline task yielded a graded acceptability judgment for each sentence pair (on a scale from 0-5, ‘totally unacceptable’ ↔ ‘perfectly acceptable’); here, the data analysed were the *aggregate scores* assigned to each pair of test items.

Notice that in neither task were subjects asked to make *explicit* metalinguistic judgments about grammaticality. Instead, the task provided a

more implicit measure. The test pairs — actually quadruplets of pairs, see above — were so constructed that only formal grammatical properties distinguished the parallel from the non-parallel pairs: thus, to the extent that significant differences were observed between the responses in the four compared conditions, the task can be claimed to have offered a direct measure of implicit competence.

2.2.3. Predictions

Based on the results of previous experiments (Tanenhaus & Carlson (1990), Duffield and Matsuo (2001)), our general prediction was that items with non-parallel antecedents (of all types) should be accepted significantly less often than corresponding items with parallel antecedents, and that those non-parallel items that *were* accepted should take longer to accept than the corresponding parallel items. Expressed in statistical terms, we predicted a main effect of syntactic parallelism by all three dependent measures.

Previous work also led us to predict a reliable interaction between syntactic parallelism and anaphor type: by all measures, non-parallel antecedents with VPA completions should be significantly more acceptable than those with VPE completions.

The present design yielded a number of additional predictions. Specifically, supposing that the non-structural factors discussed above contribute to the Parallelism Effect, we predicted statistically reliable interactions between syntactic parallelism on one hand, and finiteness and construction type, on the other. In addition, we expected a reliable parallelism effect for VPA, albeit one that was reduced relative to the effect for VPE. Finally, we predicted that all effects of parallelism should be gradient, rather than categorical: violations of parallelism should yield systematically *less* acceptable, as opposed to *unacceptable*, sentences.

With respect to L2 acquisition, if advanced L2 learners employ the same implicit competence mechanisms as native speakers, then we expected to find no significant interactions between the two subject groups with respect to the observed effects (except possibly for a main effect of subject group on the latency measure: native-speakers might be predicted to respond more quickly than L2 learners).

Interim Discussion: Distinguishing Sources of Gradience

Notice that the different dependent measures involved in these tasks generate distinct types of non-categorical response to violations of parallelism. In the acceptability judgment measure of the online task, subjects cannot directly rate the relative acceptability of any particular completion along a continuum: they can only accept or reject the stimulus categorically. Nevertheless, a kind of gradience emerges indirectly from the *aggregate rate of acceptance* for items belonging to particular conditions. In the ideal case for competent gradience, subjects should accept parallel utterances 100% of the time, reject control sentences 100% of the time, and accept non-parallel stimulus pairs 50% of the time. (In fact, the actual results closely approximated this ideal scenario).

With this dependent measure, however, it is important not only to consider the *mean* acceptance rate for experimental items in different conditions, but also to know the *modal* distribution of responses within a particular condition. An effect should only legitimately be treated as gradient if the distribution within the relevant condition is uni-modal, rather than bi-modal. For example — again idealising somewhat — a 50% mean acceptance rate for non-parallel completions could result *either* from all of the sentence-pairs in this condition being accepted 50% of the time by a particular subject group (uni-modal distribution), *or* from half of the items being accepted 100% of the time, with the other half being consistently rejected (a bi-modal item distribution). In the latter case, the observed gradience for this condition overall would be a statistical artifact, and would not accurately reflect what were in fact speakers' (implicit) categorical judgments of individual items.¹³

The other dependent measure in the online task, response latency, *is* a continuous one; therefore, it may seem more straightforward to interpret gradience in this case. In fact it is not, since rather than a *three-way* comparison between fully acceptable (parallel pairs), less acceptable (non-parallel pairs) and unacceptable (control pairs), with the non-parallel pairs intermediate on a scale of acceptability, the response latency measure only yields a *two-way* comparison: between the times for accepted parallel items with those of accepted non-parallel items. The reason for this is that the response latencies for rejected (control) items cannot be directly compared with the response times for the acceptances, since typically, it is easier to

¹³ In this extreme example, of course, such a polarised bi-modal distribution would probably fail to yield statistical significance in ANOVA or t-test analyses, given the amount of variance in the data. Thus, such statistical tests help to 'rein in' any tendency to infer competent gradience from the descriptive statistics alone; that is, from the raw means.

reject an item that is obviously anomalous as it is to accept a good, parallel completion; this yields faster rejections than acceptances.

Now, the overall experimental prediction about this measure — *viz.*, that parallel items should be accepted reliably more quickly than non-parallel items — is clear; it is also borne out by the results. Ironically though, the immediately applicable statistical analysis (ANOVA) does not allow any direct inference to competent gradience: in the absence of comparable response latencies for the rejected items, it can only be determined whether there is a significant difference between parallel and non-parallel items, something that would also be consistent with non-parallel items being treated as categorically unacceptable, rather than being of intermediate acceptability.¹⁴

The scalar response of the *offline* task does more transparently reflect speakers' gradient judgments about each item — supposing these are real — than do the online measures. In principle, subjects are free to rate each individual test sentence as intermediate in acceptability — say, 2 or 3, on a 0-5 scale. The fact that our subjects showed a nearly discrete, and reliable, tri-modal distribution in their responses provides nice confirmation of the reality of competent gradience.

However, an obvious difficulty in this case is that there is a higher risk in offline tasks that the judgments reflects explicit, metalinguistic knowledge: see, for discussion, Birdsong (1989), Birdsong (1992). This metalinguistic

¹⁴ Of course, one would then need to explain why these items were accepted at all in any numbers, in contrast to the unacceptable control items. The simple methodological point remains, however: it is only possible to classify a judgment as intermediate where there are data points on either side.

knowledge may yield more or less categorical results; typically the former, as metalinguistic information tends to be more proscriptive.¹⁵

The most important issue here is whether these different measures produce convergent results. If they do, and if non-parallel items emerge as consistently intermediate between fully acceptable and fully unacceptable items by all applicable measures, then it seems reasonable to claim that competent gradience is a real phenomenon, rather than a statistical or methodological artifact.

2.2.4. Results and Discussion

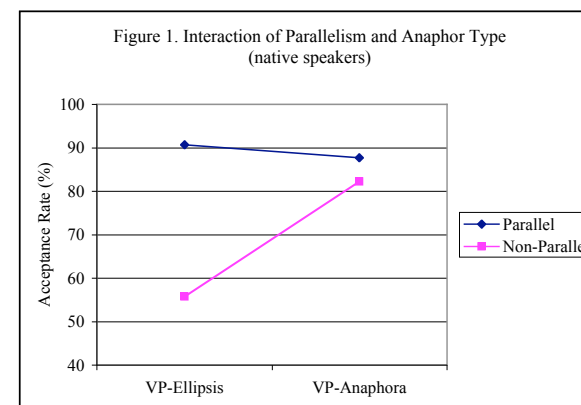
First, I present the results from native-speakers, followed by those of the second language learners. The discussion section treats both subject groups together.

2.2.4.1. Native Speakers' Results

Analyses of the *judgment* data in the online task revealed reliable main effects of syntactic parallelism, anaphor type and finiteness (with finiteness, rather than syntactic parallelism, showing the largest main effect). For the native-speakers, there was no main effect of construction type, nor any interaction between construction type and other factors.¹⁶

¹⁵ Another difficulty is that the scale is an interval scale, rather than a continuous one: while in principle, subjects could assign truly intermediate values (by circling in one of the space between the integers of the scale), in practice, subjects choose a discrete integer. Here again, intermediate values, e.g., 3.72, are a property of the aggregate scores for each condition, not of the individual scores for any given item.

¹⁶ Since this paper is concerned with a broader issue, the presentation of the results here is relatively brief. For more detailed analysis, see Duffield & Matsuo (2003a, in prep.).



As predicted, a significant two-way interaction was observed between syntactic parallelism and anaphor type, as illustrated in Fig. 1. A reliable interaction was also observed between syntactic parallelism and finiteness: as Fig. 2 below shows, the parallelism effect was significantly weaker for non-finite than for finite ellipsis. However, planned comparisons revealed significant effects of structural parallelism even in the non-finite conditions.

As for the *latency* data, separate analyses of variance revealed a reliable main effect of finiteness, as well as a marginal effect of syntactic parallelism.

The results of the *offline* experiment largely confirmed those of the online task. Separate analyses showed reliable main effects for syntactic parallelism, finiteness, and anaphor type. Although reliable two-way interactions were observed in the by-subject ANOVA — between anaphor type and syntactic parallelism, construction type and finiteness, and syntactic parallelism and finiteness, respectively — these were only marginally significant in the item analysis.

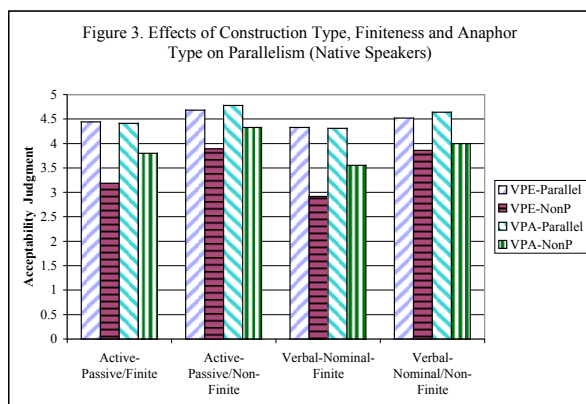
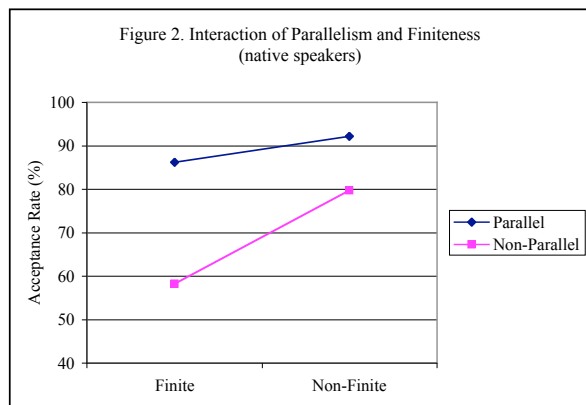


Fig. 3 above shows a breakdown of the parallelism effect by construction type, and finiteness and anaphor type in the offline task. The figure shows the striking influence of finiteness on the strength of the parallelism effect:

across both construction type and anaphor type — non-finite ellipsis elicits a significantly weaker parallelism effect than does finite ellipsis.

2.2.4.2. Second Language Learners

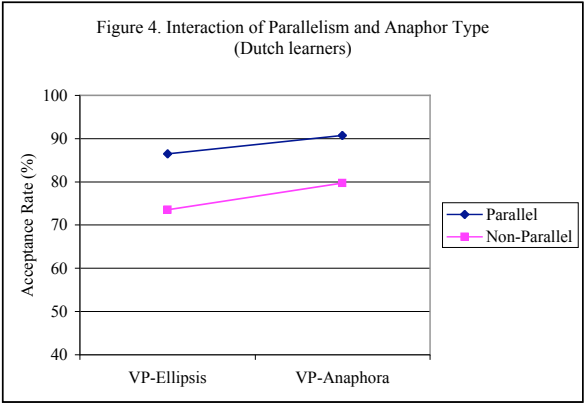
The detailed results from the *judgment data* (online task) for the Dutch subjects differ from those of the native-speakers, even though the overall parallelism effect is very similar for the two groups. The two groups diverge in terms of the main effects, with Dutch learners showing significant main effects for *all* factors, including Construction Type: in contrast to the English native-speakers, Dutch learners significantly preferred passive to nominal antecedents.

Dutch learners also diverged from native-speakers in terms of the observed interactions. The first difference is the *absence* of any interaction between Syntactic Parallelism and Anaphor Type: for Dutch learners, the strength of the Parallelism Effect in VPA contexts was not significantly different from that in VPE. This is graphically represented in Fig. 4 below (compare Fig. 1 above), where the parallel lines in the figure show that size of the effect is constant across anaphor types. Notice that in spite of this Dutch learners still differentiate the two anaphor types, accepting VPA significantly more often than VPE.

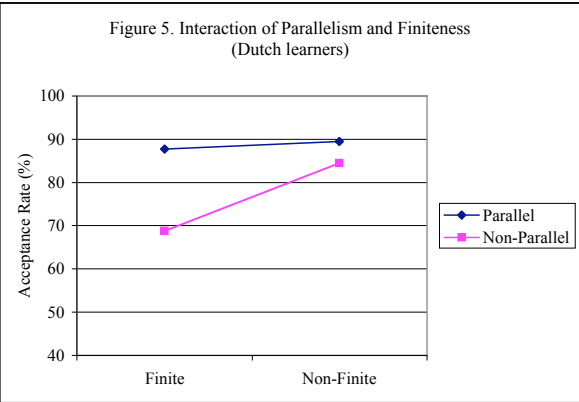
At least as important were the interactions that *did* occur in the Dutch learners' data, especially that between Syntactic Parallelism and Finiteness: for Dutch learners — again in contrast to native speakers — *the parallelism effect is virtually eliminated in non-finite contexts*.¹⁷ This is shown by the

¹⁷ Fig. 5 does show a small numerical difference between parallel and non-parallel items in the non-finite condition; however, this difference is not (even marginally) reliable statistically. Thus, this result is statistically non-distinct from a pattern of results in which the acceptance

interaction plot in Fig. 5 below. There was also a significant interaction between Syntactic Parallelism and Construction Type: the Parallelism Effect was significantly stronger with nominal, as opposed to passive, antecedents.



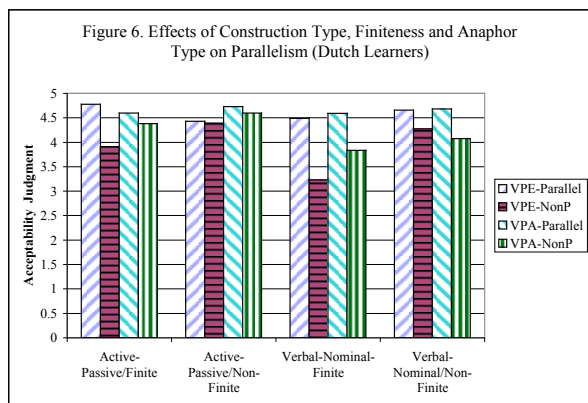
rate for parallel vs. non-parallel items was in fact completely identical, hence the expression 'virtually eliminated.'



As was true of native speakers, the Dutch learners' *latency data* were largely consistent with their judgment data, with significant main effects of Finiteness and Anaphor Type; the effect of Syntactic Parallelism was also significant by subject though not by item. There were no reliable interactions in the latency data.

As for the Dutch *offline* results, these were again wholly consistent with the online measures. Analyses of variance showed reliable main effects for all factors. In addition, two reliable two-way interactions were observed in the Dutch learners' judgment data: between Syntactic Parallelism and Construction Type, and Syntactic Parallelism and Finiteness, respectively: ellipsis sentences following (non-parallel) nominal antecedents were judged significantly less acceptable than those following (non-parallel) passive antecedents; finite ellipsis was significantly less acceptable than non-finite ellipsis following non-parallel antecedents. Finally, a significant three-way interaction was observed between Syntactic Parallelism, Anaphor Type and Finiteness. As indicated in Fig. 6, there was no significant Parallelism Effect

for either anaphor type in non-finite ellipsis contexts: only in finite contexts is a Parallelism effect observed, and only in finite contexts are responses to VPE vs. VPA completions clearly differentiated.



2.2.4.3. Discussion

The results show very close convergence between the online and offline experiments. The fact that both tasks elicit the same pattern of gradient effects and significant Parallelism Effects for VPE — as well as VPA in certain condition — strongly suggests that these are not task-specific results.

The figures also demonstrate a clear difference between English native-speakers and Dutch L2 learners with respect to the relative weighting of syntactic vs. other factors underlying the Parallelism Effect. The saw-tooth pattern in the English (VPE) results, with nearly consistent distance between the parallel and non-parallel conditions, reflects the overwhelming dominance of structural information in determining the Parallelism Effect for native-speakers. With the exception of Anaphor Type, no other factor comes close to having a significant influence in the offline experiment. The online

results are slightly more complicated, but the fundamental conclusion is the same: for native-speakers, syntax matters more than anything else.

For the Dutch L2 learners, on the other hand, *non*-structural factors, especially finiteness, play a much larger role in the Parallelism Effect (in both tasks). For L2 learners it is a crucial matter *which* non-parallel construction type forms the antecedent clause, and whether the ellipsis is finite or non-finite, since the interaction of these two factors essentially eliminates any effects of structural violations.

Thus, the results clearly suggest that the Parallelism Effect is explained by different factors for different groups of learners.

With regard to the specific issues investigated in this set of experiments, the answers vary according to which subject group is considered.

Taking first the issue of Construction Type, the native-speakers' results are consistent with the standard generative assumption that there are no construction-specific rules, since the strength of the Parallelism Effect is completely uninfluenced by the particular construction type of the carrier sentence. For L2 learners, by contrast, there is simply no evidence, *in these results*, for any construction-independent syntactic constraint. Rather, the observed effects are more consistent with an internalized competence involving highly construction-specific rules and constraints.

Similar remarks apply to Finiteness: although finiteness influences the *strength* of the effect for native-speakers — something which certainly requires an explanation¹⁸ — it makes no crucial difference to the *fact* of the parallelism effect, since there is a significant effect of parallelism in all relevant VPE conditions, irrespective of finiteness.¹⁹ Thus, while the effects

¹⁸ See Duffield and Matsuo (2003a), (in prep.) for a possible account in terms of the underspecification of non-finite clauses.

¹⁹ As revealed by planned comparisons of both online and offline judgment data.

of finiteness do speak against the assumption that the Parallelism Effect is *uniquely* due to syntactic factors, they do not directly challenge the idea that the constraint itself exerts categorical effects.

For the Dutch learners, on the other hand, finiteness — or rather, *non-finiteness* — basically eliminates the parallelism effect: in our experiment, non-finite violations of parallelism are treated as though they were grammatical. Thus, for Dutch learners, one cannot claim that the parallelism *constraint* is categorical; certainly, the *effect* is not, since it fails to generalise across conditions.

As for the effects of Anaphor Type — and the related theoretical dichotomy between surface *vs.* deep anaphora — the two subject groups again diverge. Recall that the standard theoretical analysis predicts that the parallelism effect should apply only to VPE, and should spare VPA entirely.

As predicted, English native-speakers' results show consistent, reliable interactions between Parallelism and Anaphor Type on all overall measures. Moreover, planned comparisons on the judgment data of the individual conditions reveal a reliable Parallelism Effects for all and only the *VPE* conditions. At this level of analysis, therefore, the effect of parallelism on VPA is negligible for native-speakers, at least statistically speaking.²⁰

Once again, though, the Dutch learners' results speak against conventional theoretical assumptions, since — in those conditions where it occurs at all —

²⁰ Notice the importance here of the phrase 'at this level of analysis'. In the overall analysis of VPA sentences, there is clearly an effect of syntactic parallelism. However, this effect is *not* statistically reliable, and *is* significantly smaller than the effect for VPE. Likewise, with respect to particular subconditions — e.g., VPA-finite-passive — there is no reliable effect of parallelism for VPA pairs. Thus, it is justified to claim that the effect is negligible. Yet, if there had been slightly less variance in the sample, or if more subjects had been tested, it is quite possible that some of these effects would have reached significance. This at once highlights the value, and the limitations, of a statistically-driven analysis; see also fn. 9.

the parallelism effect is just as strong for VPE as for VPA (as Fig. 4 above indicated).²¹

Hence — at this particular level of analysis — the Dutch learners' results again fail to support the theoretical assumption that parallelism applies only in contexts of surface anaphora.

In summary, with regard to the specific empirical questions raised in this experiment, the results of the more detailed analyses suggest that the conventional theoretical assumptions regarding the Parallelism Effect are basically correct for native-speakers, but wrong for second language learners (who nevertheless achieve an impressively native-like performance in terms of standard, overall, measures).

3. Interpreting Competent Gradience

Stepping back from the details of this particular experiment, we can now reconsider some of the more general conceptual and methodological issues that were introduced at the outset.

First, these results serve as an empirical reminder of the obvious — but often ignored — fact that grammaticality is not the same thing as acceptability. This much is agreed on by almost all linguists, both by those formal linguists whose impulse is to idealise away from non-grammatical aspects of acceptability judgments, and by those — typically psycholinguists — who would rather discount the notion of grammaticality as epiphenomenal.

What I tried to show in the previous paper, and in the study discussed here, is that competent gradience is a theoretically interesting phenomenon/

²¹ Notice, nevertheless, that Dutch learners *do* clearly differentiate overall between VPE and VPA (as was shown in Fig. 3): the difference, however, does not lie in the relative influence of structural parallelism but rather in the way anaphor type interacts with finiteness and construction-type.

capacity *in its own right*, and that it is possible — to some extent at least — to tease apart structural and non-structural aspects of acceptability, so as to gain a better understanding of what underlying linguistic competence is about. On this view, (final state) competence is conceived of as a much richer, more involved, set of underlying capacities than most formalists allow: in this sense, competence is much more than grammaticality. Here, the formalist view is unduly narrow.

On the other hand, it seems equally mistaken to try to eliminate autonomous grammaticality from the theoretical picture by denying the validity of the competence-performance distinction, since the results discussed here — as well as those of other experiments discussed previously — provide considerable support for both ideas. These notions require careful revision, not rejection.

These papers thus suggest a middle way in understanding and explaining judgment data. Their purpose is not to reconcile opposing ideological camps through some weak compromise, but rather to hint at some of the positive empirical results that flow from taking a moderate theoretical position.

In this regard, comparisons between native-speakers and L2 learners are especially revealing, empirically and methodologically. It is reasonable to suppose that a standard formalist analysis of L2 learners' knowledge of VPE ellipsis would conclude that Dutch L2 learners have attained the same knowledge as English speakers: after all, the Dutch learners have 'come to know' the Parallelism Effect in a non-native construction, and correctly to differentiate between VPE and VPA (without explicit instruction). Conversely, a more functionally-driven, construction-based approach to grammatical representation would also probably conclude that native speakers and L2 learners know the Parallelism Constraint in essentially the same way: if one denies the possibility of autonomous — non-construction-specific — constraints, then native speakers' acceptability judgments can be

interpreted simply as judgments on particular constructions (which is just what the L2 learners seem to be making), rather than being determined, in part, by an autonomous, categorical grammatical constraint.

In contrast to both positions, an approach that takes seriously the complexity of competent gradience, while holding open the possibility of an autonomous grammatical competence (as one component of this larger competence), is better able to address the fundamental acquisitional issue of what to make of native-speaker/L2 learner differences, as well as the more general issue of how best to understand the competence-performance distinction.

Finally, with respect to more general theories of linguistic variation, this paper offers a middle way of a different kind. Traditional approaches to variation are typically concerned with the *external* products of language: whether at a phonological, lexical or syntactic level, the focus is on charting variation in attested (produced) words or utterances across different varieties, and — more occasionally — in trying to explain the constraints on this type of variation. By contrast, recent generative approaches to variation have been concerned with *internal* variation with respect to rather narrow grammatical principles, the working assumption being either that such formal variation is extremely constrained — as in the Principles and Parameters of Government-Binding Theory (Chomsky 1981, 1986), or that it is non-existent: in Minimalism, there is no theoretically interesting variation (Chomsky (1995).

A theory of competent gradience lies intermediate between these two: altogether more psychological than traditional approaches to linguistic varieties, yet broader than in its scope than theories of narrow competence, it offers a third way, an alternative internal account of variation in language performance.

References

Allen, Joseph/Seidenberg, Mark S (1999): The emergence of grammaticality in connectionist networks. In: Macwhinney, Brian (ed.): The emergence of language, Mahwah, NJ: Erlbaum, 115-52.

Birdsong, David (1989): Metalinguistic performance and interlinguistic competence. New York: Springer-Verlag.

— (1992): Ultimate attainment in second language acquisition. *Language*, 68, 706-55.

Chomsky, Noam (1981): Lectures on Government and Binding: The Pisa Lectures. vol. 9. Dordrecht: Foris.

— (1986): Knowledge of Language: Its Nature, Origin and Use. New York: Praeger.

— (1995): The Minimalist Program. Cambridge, Massachusetts: MIT Press.

Chung, Sandra/Mccloskey, James (1983): On the Interpretation of Certain Island Facts in GPSG. *Linguistic Inquiry*, 14, 704-13.

Culicover, Peter (1998): The minimalist impulse. In: Culicover, Peter W./McNally, Louise (eds.): The limits of syntax, New York: Academic Press, 44-77.

Duffield, Nigel (2003): Measures of Competent Gradience. In: van Hout, Roeland/Hulk, Aaafke/Kuiken, Folkert /Towell, Richard (eds.): The Interface between Syntax and the Lexicon in Second Language Acquisition, Amsterdam & Philadelphia: John Benjamins Publishing Company, 97-127.

Duffield, Nigel/Matsuo, Ayumi (2001): A Comparative Study of Ellipsis and Anaphora in L2 Acquisition. In: Do, Anna H.-J. /Dominguez, Laura/Johansen, Aimee (eds.): Proceedings of the 25th Boston University Conference on Language Development, Somerville, MA: Cascadilla Press, 238-49.

— (2003a): Acquiring Competent Gradience: factoring out the Parallelism Effect in VP-ellipsis. McGill Working Papers in Linguistics, 18, 17-80.

— (2003b): Factoring out the parallelism effect in VP-ellipsis. To appear in: Cihlar, Jonathon (ed.): Proceedings of the 39th Regional Meeting of the

Chicago Linguistic Society: Main Session, Chicago: Chicago Linguistic Society.

Hankamer, Jorge/Sag, Ivan (1976): Deep and surface anaphora. *Linguistic Inquiry*, 7, 391-428.

Hardt, Daniel (1993): Verb Phrase Ellipsis: Form, Meaning and Processing, Computer and Information Science, University of Pennsylvania: Ph.D. dissertation.

Kluender, Robert (1992): Deriving island constraints from principles of predication. In: Goodluck, Helen/Rochemont, Michael (eds.): Island Constraints: Theory, Acquisition and Processing, Dordrecht & Boston: Kluwer Academic Publishers, 195-222.

Maurer, G/Tanenhaus, Michael K/Carlson, Greg N (1995): A note on parallelism effects on processing verb phrase anaphors. *Language and Cognitive Processes*, 10, 1-12.

Sag, Ivan (1976): Deletion and logical form, MIT: doctoral dissertation.

Sag, Ivan/Hankamer, Jorge (1984): Towards a theory of anaphoric processing. *Linguistics and Philosophy*, 7, 325-45.

Seidenberg, Mark S/MacDonald, Mary-Ellen C (1999): A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23, 569-88.

Tanenhaus, Michael/Carlson, Greg N. (1990): Comprehension of deep and surface verbphrase anaphors. *Language and Cognitive Processes*, 5, 257-80.