HAL
open science

Acceptable ungrammaticality in sentence matching

Nigel Duffield, Ayumi Matsuo, Leah Roberts

To cite this version:

Nigel Duffield, Ayumi Matsuo, Leah Roberts. Acceptable ungrammaticality in sentence matching. Second Language Research, 2007, 23 (2), pp.155-177. 10.1177/0267658307076544 . hal-00570725

HAL Id: hal-00570725

https://hal.science/hal-00570725v1

Submitted on 1 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Second Language Research 23,2 (2007); pp. 155–177

Acceptable ungrammaticality in sentence matching

Nigel Duffield and **Ayumi Matsuo** *University of Sheffield*
and **Leah Roberts** *Max Planck Institute for Psycholinguistics*

Received August 2005; revised December 2005; accepted January 2006

This article presents a new set of experiments using the *sentence-matching paradigm* (Forster, 1979; Freedman and Forster, 1985; see also Bley-Vroman and Masterson, 1989), investigating native speakers' and second language (L2) learners' knowledge of constraints on clitic placement in French. Our purpose is three-fold:

- to shed more light on the contrasts between native speakers and L2 learners observed in previous experiments, especially Duffield and White (1999), and Duffield *et al*. (2002);
- to address some of the specific criticisms of the sentence-matching paradigm levelled by Gass (2001); and
- to provide a firm empirical basis for follow-up experiments with L2 learners.

The results reported here provide some confirmation of the validity of Duffield *et al.*'s earlier work, and help to adjudicate among competing interpretations of the previous effects.

I Introduction

The sentence-matching task (SMT) is a relatively simple instrument: participants are asked to decide whether a pair of consecutively presented sentences are identical in form, as they are in (1a), or not, as in (1b).[1] Psycholinguistic interest in the SMT lies in the fact that in

[1]Although the SMT methodology is designed to probe grammatical knowledge, non-matching distracter pairs often differ from each other lexically, rather than grammatically; thus the pair in (1b) is typical of distracter sentences.

Address for correspondence: Nigel Duffield, Department of English Language and Linguistics, University of Sheffield, Sheffield S10 2TN, UK; email: n.g.duffield@sheffield.ac.uk

© 2007 SAGE Publications 10.1177/0267658307076544

most instances grammatical pairs are matched reliably more quickly than ungrammatical ones; for example, the sentences in (1a) are matched significantly more quickly than those in (1c) (non-matching pairs such as (1b) are simply distracters).

The explanation originally offered by Freedman & Forster for this difference is that when participants read a grammatical sentence, they construct a higher-level structural representation, which is then matched against the structural representation of the second sentence; in the case of an *un*grammatical string, no such representation can be constructed, and so participants must resort to some less efficient – perhaps word-by-word – mechanism for matching the two sentences, which results in increased response latencies for such items.

1) a. Alex has seen the dog.
   Alex has seen the dog.
   b. Alex has seen the dog.
   Alex has seen the cat.
   c. Alex has saw the dog.
   Alex has saw the dog.

Whether or not this explanation is correct is perhaps less important here than the fact that – in most cases – it works, and that it provides an implicit measure of grammaticality. The utility of such a measure has not escaped second language researchers, where explicit meta-linguistic judgements may be less likely to reflect underlying competence than is the case for native speakers (Birdsong, 1989); for this and other reasons, Bley-Vroman and Masterson (1989) propose sentence matching as a valuable adjunct to traditional grammaticality judgement (GJ) tasks in L2 research.

1 In the absence of a grammaticality effect

What is arguably of greater theoretical interest are those cases where the sentence-matching task apparently 'fails': that is, where superficially ungrammatical constructions are treated in terms of response latency as though they were grammatical. These results have been taken by some researchers as evidence of the 'underlying grammaticality' of such constructions, especially where there is independent support for this claim from syntactic theory; see, for example, Freedman and Forster (1985), Forster and Stevenson (1987), Clahsen and Hong (1995). This is also the case in the present experiment.

One such construction from French, studied in Duffield and White (1999) and Duffield *et al.* (2002), is what may be termed the '*medial causative' (*MC), where an accusative (subject) clitic is placed between the higher causative verb and the lower predicate with which it is thematically associated, as in (2a). With respect to the surface grammar, this structure is ungrammatical; however, in previous SMT experiments, *MCs were treated by native speakers no differently from their grammatical counterparts (2b), and in clear contrast to comparable ungrammatical pairs, such as those exemplified in (2c), which elicited reliably longer response latencies.

2) a. * Je fais *le* rire aujourd'hui.
I make him laugh-INF today
'I make him laugh today.'
b. Je *le* fais rire aujourd'hui.
I him make laugh-inf today
'I make him laugh today.'
c. * Je fais rire *le* aujourd'hui.
I make laugh-INF him today
'I make him laugh today.'

The absence of a grammaticality effect in this particular instance could be dismissed as a task or item effect were it not for the following considerations. First, in all other comparable conditions in the same experiment, native speakers' results yielded reliable effects of grammaticality. This was true, for example, of contrasting pairs of sentences involving so-called restructuring verbs such as *pouvoir*, *vouloir*, etc., illustrated in (3).[2]

3) a. Je veux *le* voir.
I want him-ACC see-INF.
'I want to see him.'
b. * Je *le* veux voir.

Second, the same participants' judgements of the same items in a follow-up pen-and-paper GJ task, also reported in Duffield *et al.* (2002), showed a uniform rejection of these *MC sentences, demonstrating that the absence of a grammaticality effect in the SMT was not due to poorly chosen stimuli.

[2]Note that here the expected judgements are reversed: medial placement of the clitic is grammatical, initial placement is ungrammatical.

Third, experiments using separate materials and a different instrument – a self-paced reading task – have shown independently that this construction is rather marked in processing terms. Dwivedi and Hoover (1996) and Hoover and Dwivedi (1998) have shown that readers prefer to associate a clitic with the verb that immediately follows it, a preference that yields longer processing times[3] for (grammatical) initial causatives, such as those in (2b) above and (4a), over the corresponding causatives without clitics such as those in (4b), as well as over non-causative restructuring sentences where the clitic is adjacent to the verb with which it is thematically associated, as in (5b). We return to this below.

4) a. Jean les fait entraîner (à Sara).
John them-ACC makes train-INF. (to Sarah)
'John has them trained (by Sarah)/John has Sarah train them.'
b. Jean fait entraîner les éléphants (à Sara).
John makes train-INF the elephants (to Sarah)
'John has the elephants trained (by Sarah)/John has Sarah train the elephants.'

5) a. Jean veut les entraîner.
John want them-ACC train-INF.
'John want to train them.'
b. Jean veut entraîner les éléphants.
John wants train-INF the elephants
'John wants to train the elephants.'

Fourth – the main point of interest in the earlier articles – Spanish L2 learners of French did in fact show reliable grammaticality effects for the *MC condition in the same SMT experiment; in this respect, they appeared to outperform native speakers (Duffield and White, 1999; Duffield *et al.,* 2002).

Finally, as mentioned earlier, there are theoretical grounds for considering the *MC construction to be underlyingly grammatical. Most recent generative analyses of Romance causative constructions assume that French *faire* + infinitive constructions remain fundamentally biclausal, with the lower clause containing all of the functional architecture of a normal matrix clause, including clitic projections, and where the surface (initial) position of the clitic is derived by initial merger of the accusative clitic at the lower VP followed by A′-movement to the surface position. This derivation is schematized in (6) below, with the unpronounced copy of the clitic struck through. For relevant discussion, see Guasti, 1993; 1996; Sportiche, 1996; Roberts, 1997; see also Vinka, 2002.

[3]At the point of presentation of the second verb.

6) [je le_i fais [CP … [$_{\text{ClP-Acc}}$ le_i … [$_{\text{VP}}$ pro_i rire]]]]

Depending on the particular theoretical article, the details of this analysis obviously vary. Equally clearly, this is not the only legitimate analysis of French causative constructions. However, the point is that for *MC constructions there is a class of analyses in which this particular word order is 'underlyingly grammatical', whereas for the ungrammatical sentences in (2c) and (3b) no such analysis is available: these are ungrammatical at all levels of representation.[4]

Now, supposing underlying grammaticality to be more than a pure theoretical construct – if it has some psycholinguistic reality – then it is reasonable to expect that speakers should implicitly judge these sentences as essentially grammatical – or, at the very least, as significantly more acceptable than sentences that are ungrammatical at every level of syntactic analysis. In other words, in the case at hand the underlying grammaticality of *MC sentences predicts facilitation for these items in the SMT task; in principle this can explain the absence of a grammaticality effect in this case.

2 *Gass's (2001) criticisms*

Notwithstanding the preceding discussion, the SMT has been vulnerable to criticism on various counts. Early critics, notably Crain and Fodor (1987), called for a radical reinterpretation of the original findings with

[4]This directly parallels the contrast observed in Clahsen *et al.*'s (1995) study of German word order, where *verb-final word order in matrix clauses (i) was considered as acceptable as correct verb-second order (ii), while *verb-second order in embedded clauses continued to be treated as ungrammatical (iii), eliciting reliably longer response latencies than the (surface) correct verb-final order (iv). Here too, most analyses of German verb-second treat the word order in (i) as underlyingly grammatical.

i) * Hans den Mann gesehen hat.
John the man seen has
'John has seen the man.'

ii) Hans hat den Mann gesehen hat.
John has the man seen
'John has seen the man.'

iii) * Ob Hans hat den Mann gesehen
whether John has the man seen
'whether John has seen the man.'

iv) Ob Hans den Mann gesehen hat
whether John the man seen has
'John has seen the man.'

native speakers, claiming that a low-level factor of correctability better explained the early data than any appeal to more interesting theoretical notions such as underlying grammaticality.[5] A firm riposte to Crain and Fodor's account was initially provided by Forster and Stevenson (1987), who showed the effects of correctability to be orthogonal to the effects of underlying grammaticality (which they termed overgeneration). Since that time, most experiments – including those reported here – have used the SMT to investigate sensitivity to grammatical violations in sentences where it is completely trivial to correct the ungrammatical word order; hence, where Crain and Fodor's correctability account appears much more implausible.

More recently, Gass (2001) has criticized many of the previous experiments involving L2 learners on separate grounds, both methodological and empirical. Gass's principal criticisms are based on the observed contrast between previous work and her own experiments using the SMT. Those contrasts may be briefly summarized. First, in Gass's own initial experiment, L2 learners showed no reliable grammaticality effect in any structural condition (adverb placement, subject – verb agreement, or clitic placement), nor were grammatical sentences always matched faster than ungrammatical sentences by native speakers. Second, in a follow-up experiment, intermediate learners of French did not perform significantly differently from beginning learners who – according to Gass – could not reasonably be presumed able to construct such high-level structural representations. Third, the L2 results from Gass's first experiment did not reliably correlate with the same participants' offline judgements of grammaticality (indeed, a small but reliable 'reverse grammaticality effect' was observed). Finally, Gass queried aspects of the experimental design of previous experiments, including choice of materials and timing issues.

Setting aside for a moment the non-trivial methodological issues, Gass's criticisms reduce to two essential points: that, in Gass's own studies, the SMT failed to yield any reliable grammaticality effects in

[5]Crain and Fodor's position on SM is uncompromisingly harsh: '…this overgeneration account of the matching task results is incorrect[:] the relevant difference between the two sentence types is whether the ungrammaticality tends to be spontaneously corrected by participants. This correctability account of the data is of *absolutely no theoretical interest*, but unfortunately it appears to be correct' (Crain and Fodor, 1987: 123–24; our italics).

the case of L2 learners, casting doubt on the replicability of earlier results; and that mismatches were observed between participants' results in standard offline GJ tasks vs. those on the SMT, again suggesting – to Gass, at least'that the SMT does not provide a reliable measure of grammaticality.[6]

With respect to the first point, it is clear that experimental results are only valid to the extent that they are replicable; one of the main purposes of this article is to demonstrate such replicability. However, it is equally clear that failure to replicate on any given occasion does not by itself invalidate a previous finding. Moreover, various sentence-matching results have already been successfully replicated on several occasions: for example, Clahsen and Hong (1995) and Clahsen *et al*. (1995) replicated effects for subject – verb agreement violations in German that Freedman and Forster (1985) had obtained for English, while Crain and Fodor (1987) also replicated Freedman and Forster's results, albeit with a very different interpretation.

We will not dwell here on the reasons for the failure of the SMT in Gass's experiments.[7] What is crucial to notice, however, is that in her experiments L2 learners' results globally failed to show any reliable grammaticality effect; by contrast, in Duffield *et al.* (2002), and several other articles, the SMT elicited reliable grammaticality effects for L2 learners and native speakers both overall as well as in most sub-conditions:

[6]Something that is often misunderstood in discussion is that the SMT is not, strictly speaking, an online task, since the probe position is not internal to the test stimulus; rather, participants' responses follow the presentation of the whole stimulus. In this sense, the SMT yields offline judgments just like traditional GJ tasks. What the SMT has in common with many true online experimental paradigms, and what distinguishes it from other traditional GJ tasks, is that it is a timed task; and, much more importantly, an implicit task. For further discussion, see Duffield, 2003; 2004.

[7]One difference between Gass's results and Duffield *et al.*'s is worthy of mention, however; namely, the difference in the size of overall mean response latencies. In Gass's experiments typical mean latencies exceed 2000 msec; in those of Duffield *et al.*, the mean was around 1600 msec (even before applying a 3000 msec cut-off; see footnote 9 below). This contrast is possibly due to a number of participants whose responses were relatively slow overall, skewing the facilitation for grammatical pairs, thus obscuring the grammaticality effect. Gass does not mention whether participants were included as a random factor in the analysis: the fact that all results are presented as simple *t*-tests, rather than ANOVA, suggests not. Nor, apparently, was there any normalization of individual subject means. In principle, this factor alone could be the proximate cause of the failure of the SMT to yield grammaticality effects in Gass's experiments. It should also be obvious that as mean RTs get longer, a larger difference between grammatical and ungrammatical pairs is required in order to reach significance.

it was exclusively in one sub-condition (*MC) that native speakers' results failed to yield a reliable grammaticality effect.[8]

The validity of the second point, concerning the mismatches between online and offline results in Gass's experiments, rests on the premise that offline acceptability judgments provide the most reliable test of grammaticality: what Gass terms 'our "best guess" as to the state of learners' knowledge' (Gass, 2001: fn. p. 435). Without wishing to belabour the point here, we simply reject this premise; for extended discussion, see Duffield, 2003; 2004. In fact, if it were valid, there would be no real point in using other instruments such as SM to assess grammaticality, since whenever a systematic divergence were found between results, the results of traditional GJ tasks would immediately be deferred to (as is the case in Gass's article).

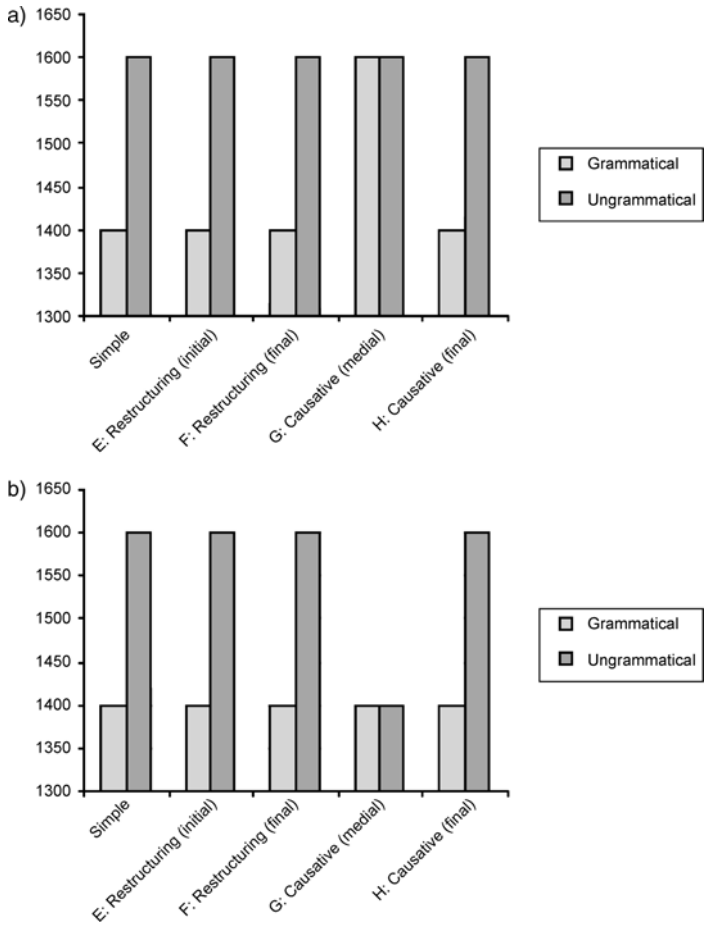
If this brief response is adequate, what remains from Gass's article is the recognition of an even greater urgency to successfully replicate the earlier results, and to develop an experimental design that allows us to better understand the grammaticality effect in the SMT. With this in mind, we conducted two follow-up experiments: the first, a near copy of the restructuring and causative sub-conditions of Duffield *et al.*'s larger experiment; the second, an extension of the causative conditions, using TRANSITIVE causatives, contrasting *medial-accusative with *medial-dative clitics. Before discussing these experiments in detail, it is necessary to address an important – but hitherto unclarified – methodological issue: namely, the locus of the (absent) effect.

3 The locus of the effect

In the SMT, the grammaticality effect (GE) is operationally defined as the difference between the mean response latency (RT or response time) for grammatical items in a particular condition, and that for the corresponding ungrammatical items (GE = $RT_{UGRAMMATICAL}$ – $RT_{GRAMMATICAL}$. In principle, the absence of such an effect in medial causatives might come about in one of two theoretically interesting ways: either, because of some processing difficulty with the grammatical

[8]Such cases of local failure are much more profitably viewed in terms of a (statistically) reliable interaction between grammaticality and sub condition than in terms of an unreliable main effect of grammaticality (as Gass would have it).

pairs, producing inhibition, or because of relatively greater ease of processing with the ungrammatical ones, producing facilitation.[9] In the ideal case, if the absence of any GE were due to inhibition, we would expect to see longer response latencies for grammatical pairs for those forms vis-à-vis other comparable grammatical pairs (see Figure 1a);



Figures 1a and 1b Contrastive effects of inhibition and facilitation (idealized)

[9]Of course, there are also completely non-interesting ways to obtain a non-result: i.e. cases where the variance across the levels of a given factor is such that there are no distinct means for grammatical vs. ungrammatical pairs.

conversely, if this were due to facilitation of the ungrammatical pairs, we would expect to see *shorter latencies* for these vis-à-vis other comparable ungrammatical pairs (see Figure 1b). Clearly, only the latter result supports the idea of underlying grammaticality. However, the results of the Hoover and Dwivedi study mentioned above, which showed readers' processing difficulties with grammatical (initial) causatives, might lead us to expect an inhibition-based result instead.

In Duffield *et al.* (2002), native speakers' results showed both effects: inhibition for the grammatical causatives (–37 msec) and facilitation for the ungrammatical medial causatives (+51 msec), when compared to the overall means (see Figure 2a). However, when the results from initial vs. *medial causatives – cf. (2a)/(2b) above, repeated here for convenience – were compared more locally with those of initial vs. *final causatives – (2b/2c) – only a strong facilitation effect (+40 msec) was found (see Figure 2b). This offers a *priori* support for the underlying grammaticality interpretation.

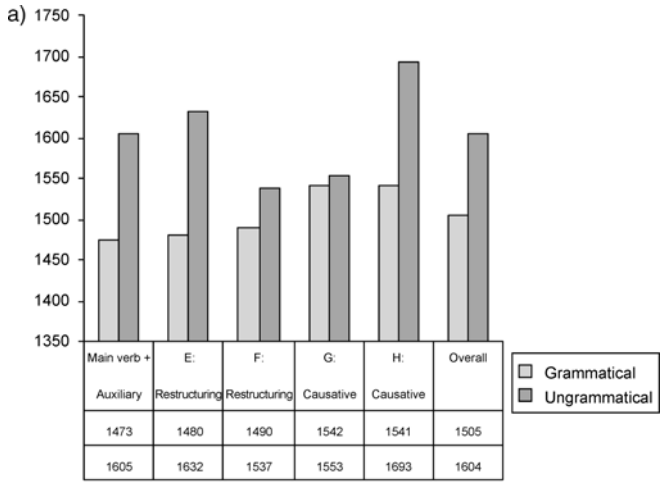
(2)	a.	* Je fais *le* rire aujourd'hui.	
		* Je fais *le* rire aujourd'hui.	1553 msec
		b. Je *le* fais rire aujourd'hui.	
		Je *le* fais rire aujourd'hui.	1542 msec
	c.	* Je fais rire *le* aujourd'hui.	
		* Je fais rire *le* aujourd'hui.	1693 msec

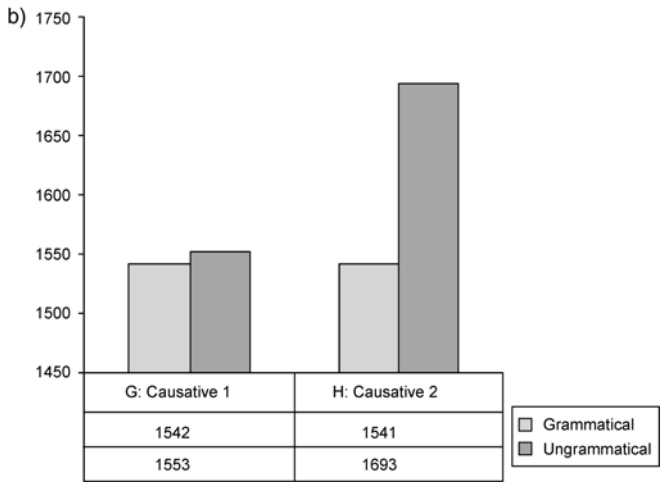
4 Section summary

In summary, our concern in this article is with the validity and replicability of the 'missing grammaticality effect' in Duffield *et al.*'s (2002) article. In light of Gass's criticisms, we are interested in demonstrating that this missing effect is a robust phenomenon, and that it occurs for the right reasons: namely, that it is due to facilitation of (surface) ungrammatical pairs in exactly one syntactic context – the *MC condition – rather than to inhibition of (surface) grammatical pairs, or to a generalized failure to elicit grammaticality effects across the board.

II Experiments

We conducted two experiments, using the same experimental set-up reported in Duffield *et al.* (2002) but with some modifications in the





Figures 2a and 2b Broad vs. narrow range comparisons

materials with respect to: the number of separate conditions in each experiment (reduced); the proportion of test items to filler/distracter items (reduced); the number of test items per condition (increased); and the number of participants tested (slightly increased). We also paid particular attention to balancing the gender/number properties of the pronominal clitics in the test items (*le, la, les lui, leur*).

1 Method (Experiments 1 and 2)

The method closely followed that of Duffield *et al.* (2002). The trials in both experiments proceeded as follows. First, a fixation point '***' was presented in the centre of the screen for 750 msec. This was followed by presentation of the first sentence towards the top of the screen (left edge: 25% vertical/25% horizontal) for 3000 msec. The first sentence then disappeared and was replaced by the second sentence of the pair, presented lower on the screen and further to the right (right edge: 75% vertical/75% horizontal). The timer was started at the onset of the second sentence. This second sentence was removed either by the participant's pressing the 'yes' or 'no' key of a button box, or by a 3000 msec time-out, if the participant failed to respond within that time.[10] The press of one or other response key (before or after the time-out) removed the second sentence and initiated the next trial, following an inter-stimulus interval (ISI) of 1000 msec.

Two versions of each experiment were prepared, one using PsyScope (Cohen *et al.*; 1993), and run on Macintosh machines, the other using NESU and run on PC workstations.[11] Experiments 1 and 2 were presented consecutively in the same testing session, always in the same order (1–2).

2 Experiment 1

a Design: Both experiments manipulated the (surface) grammaticality of the pairs, the construction type involved, the position of the clitic with respect to the two verbs (initial, medial, (final)), and the number and gender of the clitic (masc., fem.; singular, plural). Experiment 1 re-examined the contrast between causative and restructuring verbs, previously exemplified in (2) and (3) above.

b Participants: Twenty-nine adult native speakers of French recruited from two university sites in Canada and Europe participated in Experiment 1. All participants had normal hearing and normal or corrected-to-normal vision, and were paid a small fee for their participation.

[10]RTs above 3000 msec were recorded but not included in the final analysis.

[11]As well as we could determine, software/platform differences between the two versions did not materially affect the participants' perception of the task, nor did it appear to affect their performance in any way (see below).

c Materials: The materials comprised 48 (matching) test pairs and 178 pairs of practice, lead-in and distracter items (mostly non-matching distracters). The set of 48 test pairs comprised 12 items from each of 4 'condition sets': restructuring verbs with *initial vs. medial and medial vs. *final clitic placement (condition sets [E] and [F] respectively); causative verbs with initial vs. *medial and initial vs. *final clitic placement (sets [G] and [H], respectively). As in Duffield *et al.*'s experiment, the focus of interest was set G: the *MC condition. The 12 items in each condition set were themselves balanced for the gender/number of the clitic pronoun: four masculine singular clitics; four feminine singular clitics, and four plural clitics. The four experimental conditions are illustrated in Table 1.

The stimuli were organized in two versions, and ordered in such a way that each participant saw the 48 test items, 24 grammatical and 24 ungrammatical pairs. Presenting two versions of the experiment allowed us to compare directly the grammatical and ungrammatical forms of each test pair, while ensuring that no participant saw no more than one version of each stimulus type.[12]

d Predictions: Given the results reported in Duffield *et al.* (2002), we expected no difference in RT patterns according to clitic position in the restructuring verb conditions [E] and [F]: that is to say, we expected to find a grammaticality effect for this verb construction irrespective of

Table 1 Experiment 1 sample stimuli

Verb type	Set	Clitic position	Example sentence
Restructuring V	E	* initial	* Il la va manger avec ses parents.
		medial	Il va la manger avec ses parents.
	F	medial	Il va la manger avec ses parents
		* final	* Il va manger la avec ses parents.
Causative V	G	initial	Jean la fait sortir de la maison.
		* medial	* Jean fait la sortir de la maison.
	H	initial	Jean la fait sortir de la maison.
		* final	Jean fait sortir la de la maison.

[12]For example, participant one might have been exposed to pair (2a), while participant two was exposed to pair (2b): these participants would then receive the grammatical and ungrammatical tokens, respectively, of a different causative/masc.sg set of pairs from this condition. This is the type of counter-balanced list design recommended in Raaijmakers (1999; 2003), as a means of controlling item variability and thus mitigating the influence of the 'language-as-fixed-effect fallacy' (Clarke, 1973).

clitic position. By contrast, in the two causative conditions [G] and [H], we predicted an interaction between clitic position and grammaticality, with a grammaticality effect only in condition [H] (initial vs. *final): no grammaticality effect was predicted in condition set G, where the ungrammatical sentences contained the clitic in verb-medial position.

e Results and analysis: One participant correctly matched the experimental items only 48% of the time; this was 2 standard deviations below the mean of the group (94.7%), and therefore this participant's data were removed from subsequent analyses, since it was not clear that she had paid attention during the experiment. Following the removal of these data, the mean accuracy score for the group rose to 96.4% (SD: 5.6). As noted above, response latencies exceeding the 3000 msec time-out were removed, affecting 1.6% of the data. Following this, individual response latencies beyond two standard deviations of the individual mean per condition were also removed, which affected 3.02% of the remaining data. Figure 3 shows the mean response latencies per condition set for the restructuring verb and causative verb conditions in Experiment 1.

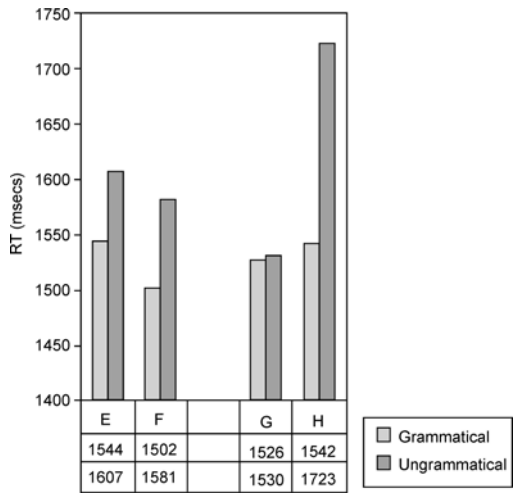


Figure 3 Experiment 1 mean latencies by condition set

f Analyses: Given the structural difference between the two verb types, two separate ANOVAs were run on the RT data to test the predictions above.

- Restructuring verb conditions (E vs. F). An ANOVA with the factors Position, (2 levels: medial/*initial vs. medial/*final) and Grammaticality (2 levels: grammatical vs. ungrammatical) was run on the RT data. As predicted, there was a significant main effect of Grammaticality, marginal by items ($F1$ (1, 27) = 5.600; $p = 0.025$; $F2$ (1, 11) = 3.421; $p = 0.091$), but no effect of Position, nor any interaction between Position and Grammaticality ($F1$, $p \geq 0.7$; $F2$, $p \geq 0.5$): in both condition sets, the grammatical pairs were responded to significantly faster than the ungrammatical pairs ($t1$ (27) = –2.366; $p = 0.025$; $t2$ (11) = –1.850; $p = 0.091$).
- Causative verb conditions (G vs. H). The same type of ANOVA was run on these data, with the factors Position (initial/*medial vs. initial/*final) and Grammaticality (grammatical vs. ungrammatical). There was both a main effect of Position in the by-participant ANOVA ($F1$ (1, 27) = 12.806; $p = 0.001$; $F2$ (1, 11) = 3.231; $p = 0.100$) and a main effect of Grammaticality, reliable participants by and by items ($F1$ (1, 27) = 9.937; $p = 0.004$; $F2$ (1, 11) = 7.212; $p = 0.021$). The predicted interaction between the two factors, which was significant by participants ($F1$ (1, 27) = 11.552; $p = 0.002$; $F2$ (1, 11) = 2.269; $p = 0.160$), was also observed.
- Paired-samples t-tests found a grammaticality effect in condition [H], where the grammatical items were responded to significantly faster than the ungrammatical items ($t1$ (27) = –4.515; $p \leq 0.001$; $t2$ (11) = –2.390; $p = 0.036$), but no grammaticality effect in condition [G], where the RTs elicited by *medial causatives were no slower than their grammatical counterparts ($t1$, $p \geq 0.9$; $t2$ $p \geq 0.7$).

In addition, there was a reliable difference between the *MC items in condition [G] and the ungrammatical causative in final position in condition [H], with the former eliciting much shorter RTs than the latter ($t1$ (27) = –4.941; $p \leq 0.001$; $t2$ (11) = –3.005; $p = 0.015$). By contrast, no difference was observed between the two grammatical conditions in [G] and [H] ($t1$, $p \geq 0.9$; $t2$, $p \geq 0.6$). This strongly supports the

prediction of facilitation for *MC constructions over other types of ungrammatical sentence.

g Interim summary: In summary, a reliable effect of grammaticality was found in all conditions except condition set G, where no GE was observed. Moreover, the ungrammatical *MC items in this condition elicited RTs that were distinctly faster than ungrammatical pairs in other conditions, and much more similar to those of the grammatical pairs in this as well as in other conditions. This clearly indicates that the absence of an effect here is due to facilitation of the (surface) ungrammatical pairs (rather than inhibition of the grammatical pairs). In short, the results of Experiment 1 were entirely consistent with our immediate predictions, as well as with Duffield *et al.*'s earlier findings.

3 Experiment 2

Experiment 2 was intended as a preliminary follow-up study, probing a contrast *within* the class of causative constructions. In contrast to Experiment 1, which involved complex sentences with an intransitive lower predicate, Experiment 2 examined the acceptability of French clitic placement with two classes of transitive causatives, contrasting the *faire* + infinitive construction with the alternative *faire par* construction. Though converging with respect to the grammatical (initial) placement of the clitic, these two constructions are distinguished by the morphological case and understood grammatical relation of the clitic pronoun. In the *faire par* construction, the clitic is interpreted as the object of the lower clause and realized with accusative case (the same case as the subject pronouns of the intransitive lower verbs discussed earlier), as in (7):

7) a. Je *la* fais suivre par son frère.
I her-ACC make follow by her brother
'I make her brother follow her.'
b. * Je fais *la* suivre par son frère.

By contrast, in the transitive *faire* + infinitive construction, the clitic is interpreted as the lower clause subject but is realized with dative, rather than accusative, case:

8) a. On *lui* fait regarder la télévision.
one her-DAT makes watch the television
'She is made to watch television.'
b. * On fait *lui* regarder la télévision.

Investigating transitive causatives enabled us to tease apart two possible interpretations of the facilitation effect for intransitive *MCs: a grammatical account vs. a processing account. The grammatical interpretation of the *MC effect – the one we have been developing up to now – predicts that sentences containing *medial accusative clitics should be treated as underlyingly grammatical independently of the grammatical relation (subject/object) of the clitic.

However, a plausible alternative to this grammatical explanation of intransitive *MC sentences is the 'shallow syntactic processing' account – along the lines proposed in Clahsen and Felser (forthcoming). This allows listeners/readers to assign a linear thematic parse to an input string, rather than constructing a detailed syntactic representation. Clahsen and Felser propose this shallow processing as a strategy predominantly employed by L2 learners (possibly to the exclusion of any 'deeper parse'). However, there is no reason why native speakers should not also make use of such a canonical thematic parsing strategy. If the absence of a GE in condition G (in Experiment 1) were due to this strategy being applied to intransitive lower subjects, then this would seem to predict that medial clitic placement would be accepted only where the ungrammatical order is congruent with canonical thematic relations (SUBJ.(AGENT)$_1$–VERB$_1$–**SUBJ.(AGENT)$_2$**–VERB$_2$): that is to say, it predicts that (2b) and (8b) pairs should pattern together, with (7b) pairs – where the clitic represents a lower clause object – showing a significant GE.

In light of our previous results however, we expected that if either of transitive constructions elicited a grammaticality effect, then it should be the dative causative (condition set [J]).[13] Thus, the two accounts of the facilitation of intransitive *MC constructions make opposing predictions for transitive *MCs.

[13]We did not exclude the possibility that condition set J might also fail to show any GE. Rather, we were agnostic here, in the absence of any theoretical account of the *faire* + infinitive construction that posits a lower (medial) position for dative clitics as well.

a Method: The procedure for Experiment 2 was identical to that of Experiment 1.

b Participants: Twenty-eight native French speakers, recruited from two University populations in Canada and Europe participated in Experiment 2. All had normal hearing and normal or corrected-to-normal vision, and were paid a small fee for their participation.[14]

c Materials: Twelve transitive *faire par* and 12 *faire* + infinitive pairs were constructed, and the clitic was placed medially to create the ungrammatical versions of these constructions. The experimental materials are illustrated in Table 2. The 24 experimental items were set within 48 fillers, and distributed evenly across two experimental versions, such that all participants saw 6 grammatical and 6 ungrammatical items of each experimental condition, but no participant saw the same item more than once.

d Results and analysis: The same data trimming procedure that was used in Experiment 1 was adopted for Experiment 2. The data from one participant who achieved an accuracy score (37.5%) which fell below 2 SDs of the group mean (94.6%) were removed from subsequent analyses. Following the removal of this participant's data, the mean accuracy score of the group rose to 96.8% (SD: 5.08). There were no individual data points above the time-out of 3000 msec, and the procedure employed to identify outliers affected 2.81% of the total data. In short, participants performed both experiments very well overall, and very few data points were lost through trimming. Figure 4 shows the mean RTs elicited in the two condition sets.

Table 2 Sample stimuli experiment 2

Verb (clitic)	Set	Clitic position	Example sentence
faire par (ACC)	I	initial	Elle les fait dessiner par les enfants.
		* medial	* Elle fait les dessiner par les enfants.
faire + inf (DAT)	J	initial	Je lui fais prendre de l'alcool.
		* medial	* Je fais lui prendre de l'alcool.

[14]These were a subset of the same participants as took part in Experiment 1. The two experiments were run consecutively in the same testing session. Due to technical problems, the data for one participant in Experiment 2 failed to be recorded (hence 29 vs. 28 participants).

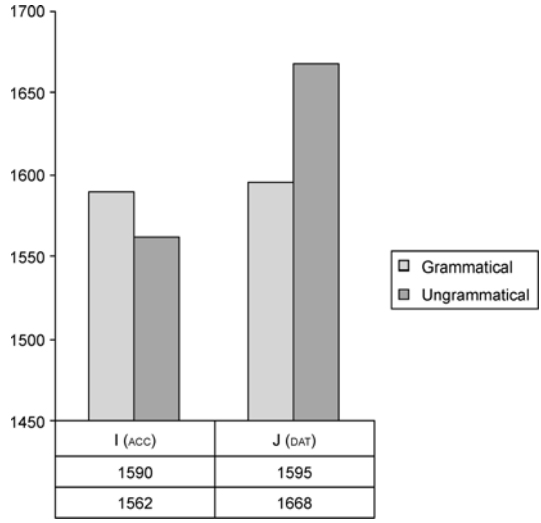


Figure 4 Experiment 2 mean latencies by condition set

e Analysis: An ANOVA with the factors Clitic Type (2 levels: accusative vs. dative) and Grammaticality (2 levels: grammatical vs. ungrammatical) was run on the RT data. There was a significant main effect of Clitic Type by participants ($F1$ (1, 26) = 4.363; $p = 0.047$; $F2$ (1, 11) = 1.409; $p = 0.260$) and marginally significant interaction between Clitic Type and Grammaticality by items ($F1$ (1, 26) = 2.623; $p = 0.117$; $F2$ (1, 11) = 3.553; $p = 0.086$). *T*-tests on the individual conditions sets showed a reliable effects of grammaticality for condition set J: the *faire* + infinitive construction with dative clitic ($t1$ (26) = –2.060; $p = 0.05$; $t2$ (11) = –2.356; $p = 0.038$). By contrast, no effect of grammaticality was observed in the accusative constructions in condition set I ($t1$, $p \geq 0.4$; $t2$, $p \geq 0.5$). Comparing across condition sets, the RTs to the ungrammatical accusative constructions differed significantly from those to the ungrammatical dative constructions ($t1$ (26) = –3.004; $p = 0.006$; $t2$ (11) = –1.955; $p = 0.077$), whereas there was no difference between the two grammatical conditions ($t1$, $p \geq 0.9$; $t2$, $p \geq 0.9$). In other words, the ungrammatical items in condition set 1 were facilitated, patterning with grammatical items in both conditions.

In brief, the results of Experiment 2 are entirely consistent with our immediate predictions, as well as with the results from Experiment 1. First, a reliable GE was observed just in the condition [J] where we had no reason to predict underlying grammaticality. Second, the absence of an effect of grammaticality in condition set 1 – as was the case in condition set G of Experiment 1 – was shown to be due to a facilitation effect for the (surface ungrammatical) items in that condition set. Thus, the results of Experiment 2 not only fully bear out those of Experiment 1, but also permit arbitration between a grammatical vs. a thematic parsing account of the facilitation effect, endorsing the former.

III Conclusions

In conclusion, the results from the two experiments reported here provide additional empirical support for the interpretation of previous results in Duffield *et al.* (2002) in terms of underlying grammaticality: the present results indicate clearly that the absence of a grammaticality effect in just these conditions is due to facilitation of the *MC items – as predicted by the underlying grammaticality hypothesis – rather than to inhibition of the corresponding grammatical controls. We also extended the domain of inquiry to transitive causatives: assuming that the contrasts observed here between dative and accusative clitic constructions are replicated in future experiments, these results provide a useful means of arbitration between different theoretical interpretations, favouring analyses couched in terms of grammatical representations over those that appeal to surface thematic processing.

Clearly, there are limits on what can be concluded from the results of two small-scale experiments focusing on one specific French construction. At least for this particular area of French syntax, however, the successful replication of the earlier experiments provides further validation of the SMT as an implicit measure of grammaticality. More importantly perhaps, these results reveal an aspect of the grammaticality of causative constructions that is not disclosed by traditional GJ tasks (whatever their other virtues may be), suggesting that in some respects at least, the SMT may be superior to those traditional tasks.

Simultaneously, the results throw into relief an otherwise unremarked contrast between native speakers, who reliably 'fail' to show a surface grammaticality effect here, and advanced L2 learners of French, who – paradoxically – 'did better' than native speakers in the earlier experiments, by treating *MC sentences as ungrammatical. This mismatch apparently in favour of L2 learners raises intriguing questions about the nature of grammatical competence, as well as the extent of ultimate attainment in second language acquisition. Although a discussion of these issues is beyond the scope of this article – see Duffield (2003; 2004) for discussion – the point to observe is that this perspective on such issues is only made possible by the use of methodologies like the SMT, which elicit implicit judgments: in this sense, the SMT more than merely serves its purpose as a 'supplement to grammaticality judgments' (Bley-Vroman and Masterson, 1989).

Acknowledgements

This article has benefited from the comments and suggestions of two anonymous *Second Language Research* reviewers, from comments from audience members at Psycholinguistic Shorts (Ottawa, 2005), and at EUROSLA (Dubrovnik, 2005), as well at colloquia presentations in various places, including University of California, San Diego, and the universities of Essex (2004), York (2005) and Edinburgh (2005). For specific comments, we are particularly grateful to Harald Clahsen and Roger Hawkins. The research reported here was supported in part by a research grant from the Social Sciences and Humanities Research Council of Canada (410–2003–0420) to Ayumi Matsuo (PI) and Nigel Duffield. It could not have been undertaken without the continuing generosity and support of Wolfgang Klein (Max Planck Institute, Nijmegen) and Juana Liceras (Ottawa), who kindly allowed us use of their lab facilities, nor without the invaluable assistance of Danijela Stojanovic (Ottawa), who ran all of the Canadian participants.

IV References

Birdsong, D. 1989: *Metalinguistic performance and interlinguistic competence*. New York: Springer-Verlag.

Bley-Vroman, R. and **Masterson, D.** 1989: Reaction time as a supplement to grammaticality judgements in the investigation of second language competence. Unpublished working papers in ESL, University of Hawai'i, Honolulu, HI, 207–37.

Clahsen, H. and **Hong, U.** 1995: Agreement and null subjects in German L2 development: new evidence from reaction-time experiments. *Second Language Research* 11, 57–87.

Clahsen, H., Hong, U. and **Sonnenstuhl-Henning, I.** 1995: Grammatical constraints in syntactic processing: sentence-matching experiments in German. *The Linguistic Review*.

Clahsen, H. and **Felser, C.** forthcoming: Grammatical processing in language learners. *Journal of Applied Psycholinguistics*.

Clarke, H.H. 1973: The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 12, 335–59.

Cohen, J.D., MacWhinney, B., Flatt, M. and **Provost, J.** 1993: PsyScope: a new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments and Computers* 25, 257–71.

Crain, S. and **Fodor, J.D.** 1987: Sentence matching and overgeneration. *Cognition* 26, 123–69.

Duffield, N. 2003: Measures of competent gradience. In van Hout, R., Hulk, A., Kuiken, F. and Towell, R., editors, *The lexicon-syntax interface in second language acquisition*. Amsterdam/Philadelphia, PA: John Benjamins.

—— 2004: Implications of competent gradience. *Moderne Sprachen* 45, 95–117.

Duffield, N. and **White, L.** 1999: Assessing L2 knowledge of Spanish clitic placement: converging methodologies. *Second Language Research* 15, 133–60.

Duffield, N., White, L., Bruhn de Garavito, J., Montrul, S. and **Prévost, P.** 2002: Clitic placement in L2 French: evidence from sentence matching. *Journal of Linguistics* 38, 1–37.

Dwivedi, V.D. and **Hoover, M.L.** 1996: Processing the French causative construction. Unpublished MS paper, McGill University, Montreal, Quebec, Canada.

Forster, K. 1979: Levels of processing and the structure of the language processor. In Cooper, W.E. and Walker, E.C.T., editors, *Sentence processing*. Hillsdale, NJ: Lawrence Erlbaum.

Forster, K.I. and **Stevenson, B.** 1987: Sentence-matching and well-formedness. *Cognition* 26, 171–86.

Freedman, S.E. and **Forster, K.I.** 1985: The psychological status of overgenerated sentences. *Cognition* 19, 101–31.

Gass, S.M. 2001: Sentence matching: a re-examination. *Second Language Research* 17, 421–41.

Guasti, M.-T. 1993: *Causative and perception verbs*. Turin: Rosenberg and Sellier.

—— 1996: Semantic restrictions in romance causatives and the incorporation approach. *Linguistic Inquiry* 27, 294–312.

Hoover, M.L. and **Dwivedi, V.D.** 1998: Syntactic processing by skilled bilinguals. *Language Learning* 48, 1–29.

Raaijmakers, J.G.W. 1999: How to deal with 'the-language-as-fixed-effect fallacy': common misconceptions and alternative solutions. *Journal of Memory and Language* 41, 416–26.

—— 2003: Another look at the 'language-as-fixed-effect fallacy'. *Canadian Journal of Experimental Psychology* 57, 141–51.

Roberts, I. 1997: Restructuring, head movement and locality. *Linguistic Inquiry* 28, 423–60.

Sportiche, D. 1996: Clitic constructions. In Rooryck, J. and Zaring, L., editors, *Phrase structure and the lexicon*. Dordrecht: Kluwer.

Vinka, M. 2002: Causativization in North Sami, Department of Linguistics. Unpublished doctoral dissertation, McGill University, Montreal, Quebec, Canada.