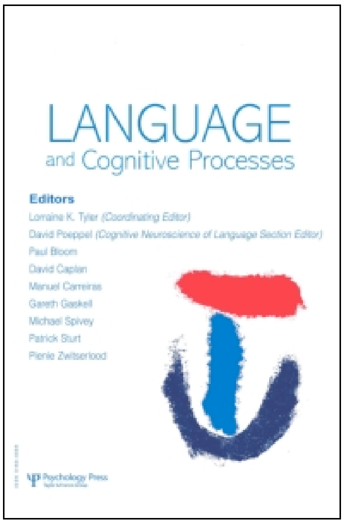

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683153

Parallel grammatical encoding in sentence production: Evidence from syntactic blends

Elizabeth Coppock [a]
[a] Stanford University, Stanford, CA, USA

First Published on: 28 May 2009

LANGUAGE AND COGNITIVE PROCESSES
0000, 00 (00), 1–12

Psychology Press
Taylor & Francis Group

Parallel grammatical encoding in sentence production: Evidence from syntactic blends

Elizabeth Coppock
Stanford University, Stanford, CA, USA

Using a large, newly available corpus of spontaneously uttered syntactic blends (e.g., *cast into question* from the targets *call into question* and *cast into doubt*) and a new method of speech error analysis, two hypotheses regarding grammatical encoding are compared: the *single-buffer hypothesis*, according to which alternative formulations of the message are encoded in the same memory buffer, potentially sharing representations, and the *multiple-buffer hypothesis*, according to which alternative formulations are independently grammatically encoded in separate buffers. Randomly generated, unattested blends were found to be reliably distinguishable from blends attested in the corpus, based on the degree to which they adhere to syntactic *alignment constraints*, controlling for other important factors. This main finding suggests that elements in similar syntactic positions across plans compete for the same slot, supporting the single-buffer hypothesis.

Keywords: Corpora; Sentence production; Speech errors; Syntactic blends.

INTRODUCTION

Syntactic blends like (1) and (2) reflect alternative formulations of the speaker's message that are 'intertwined in speech' (Garrett, 1980), shown in (a) and (b).

(1) sudden quicks ... stops aren't so bad.
  a. sudden stops
  b. quick stops

Correspondence should be addressed to Elizabeth Coppock, Stanford University, Margaret Jacks Hall, Building 460, Stanford, CA 94305, USA. E-mail: coppock@stanford.edu

Thanks to Dan Jurafsky, Carson Schütze, Victor Ferreira, Stephanie Shih, Laura Staum, and the participants of the 2005 Linguistic Society of America Summer Institute Workshop on Slips of the Tongue for their feedback when this work was presented there.

http://www.psypress.com/lcp DOI: 10.1080/01690960902840261

(2) Ed Borkowski's not letting the pressure on here.
a. . . . not letting the pressure up
b. . . . keeping the pressure on

In Garrett's (1980) words, 'the message being communicated is constant but its expression is divergent'. The present research investigates the degree to which these divergent expressions share a common representation, bringing a large corpus of syntactic blends to bear on the question (Cohen, 1987).

What does it mean for the expression of a message to be divergent? Consider the following example:

(3) It's not so pretty bad.
a. It's not so bad.
b. It's pretty good.

Here, the two alternative formulations both encode the idea of a lack of extremity, through either a combination of two adverbs: *not* and *so*, or a single one: *pretty*. Thus, alternative ways of packaging the message into content words are simultaneously expressed.

One interpretation of this divergent packaging is that blends combine alternative conceptualizations of the same message. Consider the network model of lexical selection described by Bock and Levelt (1994, p. 951): Such concepts as 'not' and 'extreme' would be found in a network of conceptual nodes, and would both be linked to the lexical concept corresponding to *pretty*. Only 'extreme' would be connected to the lexical concept for *so*. Under the assumption that among the set of concepts is a unique concept for each content word (e.g., Chang, Dell, & Bock 2006), the divergent distribution of concepts among content words indicates that the alternative formulations can reflect alternative conceptualisations of the message. Empirical support for this view comes from the fact that the two formulations are not always truth-conditionally equivalent, as exemplified in (3). Something that is not so bad is not necessarily pretty good.

Whether or not alternative conceptualizations are involved, the alternative formulations constitute diverse strategies for packaging the message content. The question, then, is how to model the parallel development of these strategies. Under the *multiple-buffer* hypothesis, multiple grammatical encoders work independently in parallel in separate memory buffers; the representations they construct do not share any representational units. Under this hypothesis, blending occurs at the articulatory level. Under the *single-buffer* hypothesis, there is only one grammatical encoder; its conceptual inputs are formulated in the same memory buffer, potentially sharing representational units. Under this hypothesis, blending occurs during grammatical encoding.

Evidence for the multiple-buffer hypothesis

One argument that could be made for the multiple-buffer hypothesis is that blends tend to take the form either of *substitutions*, the replacement of a single word in one target by a single word in another target, or *splices*, the concatenation of an initial substring from one target with a final substring from another (Fay, 1982). Most blends are describable as either type; in Fay's (1982) classification, these are *indeterminate* blends. Those that can be described as neither type are called *complex*. In Fay's (1982) corpus of 64 blends, approximately 62% are indeterminate, 12% more can be described as splices, and 13% more as substitutions. Only 5% of the corpus cannot be described in one or both of these ways. Similar numbers hold for the Cohen (1987) corpus; only 5% of the corpus can be classified as complex. The rare complex blends could simply be put aside; then, two basic mechanisms for producing syntactic blends could be posited: lexical substitution and articulatory switching from one fully specified plan to another. This would immediately derive the predominance of splice blends. However, the predominance of splices and substitutions could also be explained under the single-buffer hypothesis, assuming a preference for keeping compatible units together.

Fay (1982) describes an articulatory switching mechanism for splice blends, and argues furthermore that it is determined by prosody 'rather than' syntax, citing the following two examples:

(4) It doesn't make any matter.
 a. It doesn't make any <u>difference</u>.
 b. It doesn't <u>matter</u>.

(5) If you're not watch out ...
 a. If you're not <u>careful</u> ...
 b. If you don't <u>watch out</u> ...

In (4), the interacting units (underlined) are *matter* and *difference*; in (5), *careful* and *watch out*. These are prosodically, not syntactically, parallel units.

Fay's 'algorithm' for splice-formation goes as follows: First, mark the words in each target that bear primary and secondary stress. Then align the sentences so that primary stressed words occur in the same position, and likewise for secondary stress. For example, the alignment between the targets *What the hell am I doing here?* and *What the hell am I here for?* (which produced *What the hell am I doing here for?*) would be as in Figure 1.

Given this alignment, two types of blends may occur: A *pre-stress splice* occurs immediately before an aligned pair of stresses (as in *It doesn't make any | matter*). In a *post-stress splice*, the following unstressed words from

2		2			1	
What	**the**	**hell**	**am**	**I**	**here**	**for?**
A_1	A_2	A_3	A_4	A_5	A_6	A_7
2		2			1	
What	**the**	**hell**	**am**	**I**	**doing**	**here?**
B_1	B_2	B_3	B_4	B_5	B_6	B_7

Figure 1. Prosodic alignment between *What the hell am I doing here?* and *What the hell am I here for?* adapted from Fay (1982).

both targets are incorporated into the error (as in *What the hell am I doing here | for?).* In both types, aligned words occupy the same slot, so they never co-occur in the output.

An additional kind of support for the multiple-buffers hypothesis comes from Butterworth's (1982) argument that the "splice point" is determined by phonological similarity. As an example of this, he cites *Die Studenten haben demonstrart*, a blend of *Die Studenten haben demonstriert* 'the students demonstrated' and *Die Studenten haben demonstrationen gemacht* 'the students made demonstrations', from Meringer and Mayer's (1895) corpus of speech errors in German. The phonological similarity between *demonstriert* and *demonstrationem* is taken to explain why the splice point occurs within *demonstriert*. Assuming a modular model in which phonological information is not accessible to "higher" levels of processing, such phonological influences would suggest that syntactic blending occurs after phonological content is specified.[1] However, the idea that phonological similarity influences the location of the splice point remains to be statistically evaluated.

Evidence for the single-buffer hypothesis

One piece of evidence for the idea that alternative formulations of a message share representational units is that lemmas from one target appear sometimes to undergo morphological adjustments in order to match the syntactic requirements of the other target, as (6) and (7).

(6) He will <u>appear</u> to have the first down.
 a. He <u>appears</u> to have the first down.
 b. He will likely have the first down. (or 'It appears that he will have…')

[1] Another interpretation of this putative state of affairs is taken by Harley (1984), who rejects the modularity hypothesis on such grounds. However, the most conservative theoretical interpretation of this kind of influence would be that syntactic blends take place at the articulatory level.

(7) He will never continue to amaze me.
a. He will never cease to amaze me.
b. He continues to amaze me.

More dramatically, (8) involves an "adjustment" that deletes the subject of the small clause complement of *want* (that is, *me*):

(8) I'd be happy to help in any way I want to.
a. I'd be happy to help in any way I can.
b. I'd be happy to help in any way you want me to.

This type of blend cannot be explained by articulatory switching from one fully specified plan to another.

Moreover, if the syntactic formulation of alternative messages takes place in a single buffer, then words in similar syntactic positions should compete for the same slot. To identify competition among members of a given word class, the two targets can be aligned based on similarity, which yields correspondences across plans. Corresponding (aligned) words compete for the same slot.

Let us define *alignment constraints* to be of the form: A_i:B_j, where A and B represent the targets as word strings, and *i* and *j* are indexes into those strings. Since an alignment constraint represents a competition between the two words for the same slot, an alignment constraint is violated if both surface. Generally, an alignment constraint is violated if there is no word in the output string that the pair of aligned words could both correspond to. Formally, an alignment constraint A_i:B_j is violated if it is impossible for there to be a *k* such that A_i:O_k and B_j:O_k, where O represents the output string. (See Coppock, 2006, Appendix A for a logic and an algorithm for proving whether an alignment constraint is violated.)

In some cases, it appears that syntactically parallel units align, *rather than* prosodically parallel ones, as in the following example:

(9) If we were caught out. . . .
a. If we were caught. . . .
b. If we were found out. . . .

In (9), *caught* is syntactically parallel to *found*. If words in syntactically similar positions align, then *caught* and *found* should align, as illustrated in Figure 2, which shows the alignment constraints A_1:B_1 (*if:if*), A_2:B_2 (*we:we*) A_3:B_3 (*were:were*) and A_4:B_4 (*caught:found*). Indeed, these alignment

IN	PRP	VB	VBN	
If	**we**	**were**	**caught**	
A_1	A_2	A_3	A_4	
IN	PRP	VB	VBN	RP
If	**we**	**were**	**found**	**out**
B_1	B_2	B_3	B_4	B_5

Figure 2. Syntactic alignment between *If we were caught* and *If we were found out*.

constraints are satisfied in the attested blend; only one of *caught* and *found* surfaces.

A prosodic alignment model such as Fay's predicts that in a blend of a verb with a verb followed by a particle (e.g., *caught* plus *found out*), the single verb should align with the particle (yielding *found caught*), because they are prosodically parallel.[2] This is illustrated in Figure 3, showing the alignment constraints A_2:B_2 (*we:we*) and A_4:B_5 (*caught:out*). The constraint A_4:B_5 is violated in the blend *If we were caught out* because A_4:O_4 (*caught:caught*) and B_5:O_5 (*out:out*). The set of such blends in the Cohen corpus is in Table 1. The verb always aligns with the verb rather than the particle, in support of syntactic alignment.

There is other, published support for syntactic alignment. Fay (1982) shows that in substitution blends, the two words always have the same syntactic category. MacKay (1973) finds that even in what Fay calls 'splice blends', syntactic category is often preserved, and when it is not, a grammatically possible sentence tends to result. Cutting and Bock (1997) found that syntactic similarity facilitates blending. Asked to produce one of the two idioms they had recently heard, speakers produced blends significantly more often when the idioms had the same syntactic structure,

	2			1
If	**we**	**were**		**caught**
A_1	A_2	A_3		A_4
	2			1
If	**we**	**were**	**found**	**out**
B_1	B_2	B_3	B_4	B_5

Figure 3. Prosodic alignment between *If we were caught* and *If we were found out*.

[2] Alternatively, Fay might analyse *found out* as a single lexical unit, following his analysis of the blend *if you're not watch out* from ... *careful* and ... *watch out*, where he analyses *watch out* as a unit (p. 729). This analysis of *found out* would predict the blend to be identical to one of the targets, though.

TABLE 1
Verb + Verb-particle blends

Example	*Target A*	*Target B*
If we were **caught out**...	caught	found out
I **checked base** with David about...	checked	touched base
It **consumes up** a lot of time on the clock.	consumes	takes up
If this **continues on**...	continues	goes on
I can't **fathom out** how it could have happened.	fathom	figure out
They might **fold under**.	fold	go under
Let's **list down** all the options.	list	write down
It doesn't **meet up** to his standards.	meet	measure up
They have **proved out** very successful for me.	proved	turned out
I hate to see it **returned back** to them.	returned	given back
He **slipped down**.	slipped	fell down
There are people **waiting by** to take your call.	waiting	standing by

controlling for semantic similarity. Syntactic similarity may also help determine form; the correspondences contributing to overall syntactic similarity may represent alignment constraints, analogously to how phonological similarity in Freudian slips not only facilitates them, but also influences their formal shape (Cutler, 1982a).

CLASSIFICATION STUDY

If words in parallel syntactic positions compete for expression, then attested blends should tend to adhere to syntactic alignment constraints. Therefore, it should be possible to distinguish arbitrary mish-mashes of the target strings from attested blends based on syntactic alignment. To test this hypothesis, a set of pseudo-blends was automatically generated by merging together substrings of the targets in the Cohen corpus, and a logistic regression study was carried out in which the dependent variable was whether or not the candidate blend is attested in the corpus.

Since adherence to alignment constraints favours outputs that are more likely to be at least locally grammatical, it is important to control for the likelihood of the sequence of part-of-speech tags represented by a given pseudoblend, as well as the likelihood of the sequence of words, by a similar token. It is also important to take into account the role of lexical alignment, since many syntactic alignment constraints hold between identical words. If A = *If we were found out* and B = *If we were caught*, there would be *lexical* alignment constraints $A_0 = B_0$, $A_1 = B_1$, and $A_2 = B_2$, requiring that A and B's *If*s, *we*s, and *were*s correspond.

409. 'carry through easily' in: 'The nomination will carry through easily.'
a. The nomination will carry easily.
b. The nomination will go through easily.
(And possible influence: carry through in e.g. They carried the supplies through the jungle.

Figure 4. Example # 409 from Cohen's (1987) corpus (used by permission).

If, controlling for these additional factors, adherence to alignment constraints remains a significant predictor of whether a given candidate blend is attested in the corpus or not, we have strong evidence for the idea that elements in parallel syntactic positions tend to compete for realisation in syntactic blends, in support of the single-buffer hypothesis.

Procedure

An electronic database of blends was created using the Cohen text through scanning and optical character recognition (see Figure 4). Part-of-speech tags were manually assigned to the targets following the Penn Treebank guidelines (Marcus, Santorini, & Marcinkiewicz, 1993).

Excluded from the analysis were examples that (i) were spoken by children or non-native speakers; (ii) involved segmentations within a word boundary (e.g., *when every time I make a speech*, from *whenever I make a speech* and *every time I make a speech*); (iii) sounded fully acceptable to me; or (iv) contained more than four words (for the sake of computation time).

For each remaining blend, a set of pseudo-blends was generated. This set contained every possible merge of every substring of the targets, allowing deletions. The only constraints on the pseudo-blends were that (i) they preserve the order of the words in the targets and (ii) they contain at least one word from each target. For example, the blend in (10) had the following associated pseudo-blends: *from now hence*, *from hence now*, *hence from now*, *from hence*, *hence from*, *now hence*, and *hence now*.[3]

(10) three years from hence
a. . . . from now
b. . . . hence

For each candidate blend, the probability of both the word sequence (NGRAM_LEX), and the part-of-speech sequence (NGRAM_POS) was estimated using a trigram model of spoken conversational English with Good-Turing discounting and Katz-backoff for smoothing (the standard model in the

[3] Note that pseudo-blends might be grammatical, and then they would not have appeared in the corpus. This is a source of error, but does not bias against the hypothesis.

SRILM toolkit; Stolke, 2002). The NGRAM_LEX score was estimated from the Fisher English corpus, and the NGRAM_POS score was estimated from the POS-tagged portion of the Switchboard corpus.

Each blend was also assigned an integer representing the number of times it violated an alignment constraint, either lexical (ALIGN_LEX), or POS (ALIGN_POS). The lexical and POS alignment constraints were derived using a program called ScLITE ('Score-Lite,' from the NIST Scoring Toolkit), which is normally used for scoring the outputs of speech recognisers. ScLITE provides an alignment of two strings, minimising the number of insertions, deletions, and substitutions, and maximising the number of common elements. This program was used to align the targets, as both word and POS sequences. The common elements in the alignment between two targets represent alignment constraints. For example, the targets' POS sequences in (10) are (a) IN RB and (b) RB. Since *hence* and *now* were both tagged as adverbs (RB), they yield common elements in the ScLITE alignment, so there is a POS alignment constraint requiring that the first (only) word of target B (B_0, *hence*) corresponds to the second word of source A (A_1, *now*), i.e., $A_1 = B_0$.

For identifying violations of an alignment constraint, correspondences between the target strings and the candidate blends were identified, again using ScLITE. The ALIGN_LEX and ALIGN_POS scores equal the number of violations of the corresponding alignment constraint type.

These scores were used as predictors in a logistic regression model whose dependent variable was whether or not the blend is attested in the corpus. One pseudo-blend per attested blend was randomly chosen for analysis, to balance the sizes of the attested and non-attested classes. Random selection of a pseudo-blend and model fitting was iterated 20 times; averages over 20 models are reported below.

Results and discussion

All of the predictors (NGRAM_LEX, NGRAM_POS, ALIGN_LEX, and ALIGN_POS) significantly improved the model. The average coefficients are given in Table 2 for two models: (i) A baseline model, including just NGRAM_LEX and NGRAM_POS, and ALIGN_LEX; (ii) a full model including ALIGN_POS. The alignment constraint factors have negative coefficients, signifying that the more alignment constraint violations a blend has, the less likely it is to have been attested. Over 20 trials, these coefficients varied only negligibly ($SD < 0.05$ for sequencing constraints; $SD < 0.1$ for alignment constraints). The baseline model is significantly improved by including ALIGN_POS as a factor, by a likelihood ratio test ($\chi^2 = 404.8$; $df = 1$; $p < .001$).

Evaluation measures for each model are given in Table 3. Classification *accuracy* is the per cent of accurately classified candidates, defined based on

TABLE 2
Factor coefficients for three models

	Logistic regression model	
Factor	*Base*	*Base* + ALIGN_POS
Intercept	4.49	4.31
NGRAM_LEX	0.01	0.01
NGRAM_POS	0.04	0.04
ALIGN_LEX	−2.21	−0.68
ALIGN_POS	N/A	−1.89

positive vs. negative log odds assigned by the model. Classification *precision* is the per cent of candidates classified as attested when actually attested, and classification *recall* is the per cent of attested candidates classified as such. The *F*-score is defined as:

$$2 \times precision \times recall/(precision + recall)$$

and the R^2 value corresponds to the amount of variance explained by the model.

As Table 3 shows, most of the improvement in accuracy from adding ALIGN_POS to the model comes from precision, rather than recall. Many of the candidates incorrectly accepted by the base model were correctly rejected by the model with POS alignment (the absolute number of blends correctly accepted by the two models is relatively constant). This is due to an asymmetry in the data: identical words usually have the same part of speech. The examples correctly rejected by the full model but not by the baseline model tend to involve POS alignment constraints on distinct lexical items. For example, a pseudo-blend generated from the targets *think about* and *consider* was *think consider about*. This violates the POS alignment constraint that *think* should correspond to *consider*, whereas the attested outcome, *consider about*, does not. The full model correctly rejects *think consider about*, but the model without POS alignment incorrectly accepts it.

There were very few examples correctly accepted by the full model but incorrectly rejected by the baseline model. One of these was *by and away*

TABLE 3
Evaluation metrics for three models

Model	*Accuracy*	*Precision*	*Recall*	*F-score*	R^2
Base = NGRAM_LEX + NGRAM_POS + ALIGN_LEX	74.9	70.4	85.0	77.0	40.3
Base + ALIGN_POS	82.0	79.3	86.1	79.0	55.1

(from *by far* and *far and away*), which violates the lexical alignment constraint that *far* corresponds to *far*, but no POS alignment constraints.

CONCLUSION

Attested blends can be reliably distinguished from unattested blends on the basis of syntactic alignment, controlling for lexical alignment and word and part-of-speech sequence predictability. Additional arguments for the single-buffer hypothesis include the existence of blends involving morphological adjustments and blends that appear to involve blending at the functional level. This all supports the single-buffer hypothesis; multiple formulations of the same message are developed in the same memory buffer, and the developing syntactic representations may interact and compete with one another during grammatical encoding.

The parallel, incremental, unification-based grammatical encoding mechanism described by de Smedt (1990, 1996) seems amenable to this result, but it remains to be worked out how multiple formulations might be developed in parallel, and how phrasal units might unify despite being incompatible.

Another question to explore is the relationship between word blends and syntactic blends. There are some cases of syntactic blends in which there appear to be a break within a word (as in *When every time I...*); to what extent are these related to syntactic blends? The electronic availability of the Cohen corpus, the technology for identifying alignment constraints, and the method of comparing attested to unattested outcomes will facilitate future research into such questions.

Manuscript received November 2007
Revissed Manuscript received February 2009
First published online Month/year

REFERENCES

Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego: Academic Press.

Butterworth, B. (1982). Speech errors: Old data in search of new theories. In A. Cutler (Ed.), *Slips of the tongue and language production*. Amsterdam: Mouton.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234–272.

Cohen, G. L. (1987). *Syntactic blends in English parole*. Frankfurt: Peter Lang.

Coppock, E. (2006). Alignment in syntactic blending. In C. T. Schütze & V. S. Ferreira (Eds.), *The State of the Art in Speech Error Research: Proceedings of the LSA Institute workshop*. MIT Working Papers in Linguistics 53. Cambridge, MA: MIT Press.

Cutler, A. (1982a). The reliability of speech error data. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 7–28). Amsterdam: Mouton.

Cutler, A. (Ed.) (1982b). *Slips of the tongue and language production*. Amsterdam: Walter de Gruyter/Mouton.

Cutting, J. C. & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory and Cognition*, *25*, 57–71.

de Smedt, K. (1990). IPF: An Incremental Parallel Formulator. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation* (pp. 167–192). London: Academic Press.

de Smedt, K. (1996). Computational models of incremental grammatical encoding. In A. Dijkstra & K. de Smedt (Eds.), *Computational psycholinguistics: AI and connectionist models of human language processing* (pp. 279–307). London: Taylor & Francis.

Fay, D. (1982). Substitutions and splices: A study of sentence blends. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 163–195). Amsterdam: Mouton.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. L. Butterworth (Ed.), *Language production, Volume I: Speech and talk* (pp. 177–220). London: Academic Press.

MacKay, D. G. (1973). Complexity in output systems: Evidence from behavioral hybrids. *American Journal of Psychology*, *86*(4), 785–806.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, *19*, 313–330.

Meringer, R., & Mayer, K. (1895). *Versprechen und Verlesen*. Stuttgart: Goschensche Verlag.

Stolke, A. (2002). SRILM – an extensible language modelling toolkit. In *Proceeding of the International Conference on Spoken Language Processing*.