

# A time course analysis of enriched composition

BRIAN McELREE and LIINA PYLKKÄNEN  
New York University, New York, New York

MARTIN J. PICKERING  
University of Edinburgh, Edinburgh, Scotland

and

MATTHEW J. TRAXLER  
University of California, Davis, California

Linguistic analyses suggest that common and seemingly simple expressions, such as *began the book*, cannot be interpreted with simple compositional processes; rather, they require enriched composition to provide an interpretation, such as *began reading the book* (Jackendoff, 1997; Pustejovsky, 1995). Recent reading time studies have supported these accounts by providing evidence that these expressions are more costly to process than are minimally contrasting controls (e.g., McElree, Traxler, Pickering, Seely, & Jackendoff, 2001). We report a response signal speed-accuracy trade-off (SAT) study in which two types of expressions that are thought to require enriched composition were examined. Analyses of the full time course SAT data indicate that these expressions were interpreted less accurately and, most importantly, more slowly than control sentences. The latter finding suggests that enriched composition requires the online deployment of complex compositional operations.

Interpreting an expression requires combining the meaning of its elements according to their syntactic position. However, even seemingly simple expressions may require more complex forms of semantic composition, which Jackendoff (1997) has termed *enriched composition*. For example, simple compositional operations appear inadequate for interpreting the verb phrase (VP) in *The author began the book* (Pustejovsky, 1995). The reason is that verbs such as *begin* (*continue, finish, enjoy*, etc.) require complements that refer to events, but no simple construal of the noun phrase (NP) *the book* satisfies this requirement.

Typically, readers interpret this sentence as *The author began writing the book* or, less frequently, as . . . *began reading the book* (McElree, Traxler, Pickering, Seely, & Jackendoff, 2001). This suggests that readers “coerce” the book into the required semantic type by interpreting it as part of an unstated eventive complement (e.g., *writing the book*). Hence, these expressions can be viewed as a form of logical metonymy, because the verb’s logical structure forces an interpretation in which a part of an event, *the book*, denotes the whole event, *writing the book* (Lascarides & Copestake, 1998; Pustejovsky, 1995).

We have found that these expressions engender difficulty during reading, in agreement with the idea that they require costly compositional operations to repair the

type mismatch between the event-selecting verb and the entity-denoting object (McElree et al., 2001; Pickering, McElree, & Traxler, 2005; Traxler, Pickering, & McElree, 2002). One contrast compared coerced sentences such as (1) with controls such as (2):

1. The carpenter began the table during the morning break.
2. The carpenter built the table during the morning break.

Controls such as (2) used the verb that raters most often ascribed to the eventive interpretation of (1). Processing (1) was more costly than processing (2), even though they were rated as being equally plausible. In self-paced reading, participants took longer reading *table* and *during* in (1) than in (2) (McElree et al., 2001). In eye tracking, reliable differences first emerged at the NP (e.g., *the table*; Pickering et al., 2005) or on the two words following the NP (e.g., *during the*; Traxler et al., 2002). The same effect was found when aspectual properties were matched between the control and the coerced expressions, as when (1) was compared with the control in (3) (Pickering et al., 2005):

3. The carpenter began building the table during the morning break.

Reading patterns for (2) and (3) did not differ at *the table* or beyond, but eye-tracking measures showed that readers spent longer processing *the table* in (1) than in both (2) and (3).

Other experiments indicate that this effect lies in the act of combining an eventive verb with an entity NP. With self-paced and eye-tracking measures, Traxler et al. (2002)

---

This research was supported by National Science Foundation Grant 0236732 (awarded to B.M.). Correspondence concerning this article should be addressed to B. McElree, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003 (e-mail: brian.mcelree@nyu.edu).

found that *The boy started the puzzle* was more difficult than . . . *started the fight*, . . . *saw the fight*, or . . . *saw the puzzle*. NPs referring to entities (e.g., *the puzzle*) and those referring to events (e.g., *the fight*) did not differ as the complements of verbs such as *saw*, which can semantically combine with either argument type. However, entity NPs were more difficult to process than eventive NPs as complements of verbs such as *start*, which require eventive complements. This interaction suggests that the effect is linked to the kind of compositional operations readers deploy on the VP, and not because any complement of a verb such as *start* is difficult to interpret.

### Time Course Measures of Enriched Composition

What psychological processes are used to interpret expressions requiring coercion? Pustejovsky (1995) proposes that, formally, coercion converts an expression,  $\alpha$ , to the semantic type expected by a governing function,  $\beta$ . For example, the verb *begin* ( $\beta$ ) selects for an eventive complement, and coercion converts an NP such as *the table* ( $\alpha$ ) from its default entity semantic type to the event type. This is accomplished by selecting an activity compatible with the subject and direct object and then building an eventive interpretation, simplified here as

4. [ $\beta$ began[ $\alpha$ building the table]].

We do not attribute the coercion cost to the act of retrieving or inferring the activity implicit in the event interpretation, since coercion effects are not eliminated when the context provides the action implicit in the coerced interpretation (e.g., *The carpenter was building all morning. Before he began the table . . .*; Traxler, McElree, Williams, & Pickering, 2005). Hence, we hypothesize that the costs reflect operations readers undertake to build a representation for the event sense analogous to (4).

If readers must undertake these operations, then, crucially, coerced expressions should take longer to interpret than control expressions. Reading time patterns are consistent with this claim, but not unequivocally so. Reading time tasks provide measures of the relative difficulty of processing but do not uniquely identify the source of the difficulty. Longer reading times may indicate that readers take longer to compute a plausible interpretation, but they could also indicate that readers are less likely to accurately process all the information necessary for an interpretation (McElree, 1993; McElree & Nordlie, 1999). A case in point concerns figurative and literal expressions (e.g., *Some dancers/insects are butterflies*). Using the time course measures employed here, McElree and Nordlie found that readers compute figurative and literal interpretations with comparable speed and that longer reading times for figurative expressions, when observed, simply indicate that readers are less likely to construct a sensible figurative interpretation. We employed similar procedures here to test whether enriched interpretations require more computational time or whether readers are just less likely to compute a sensible interpretation in a given unit of time.

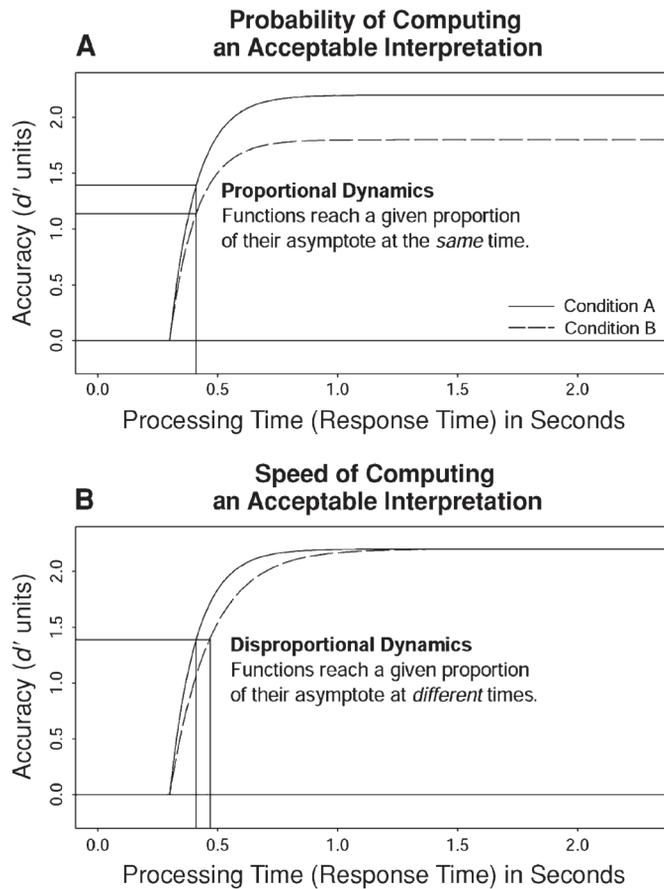
We used the multiresponse speed–accuracy trade-off (MR–SAT) procedure (Bornkessel, McElree, Schlewsky, & Friederici, 2004; McElree, 1993; Wickelgren, Corbett, & Doshier, 1980) to conjointly measure processing speed and accuracy. Participants read sentences presented phrase by phrase and decided whether they represented sensible expressions. The participants were trained to respond to a response signal (a tone) presented at 14 times following the onset of a crucial expression. The sampled times (0–4,900 msec) enabled us to fully measure how the different interpretations unfolded over time.

Figure 1 presents illustrative SAT functions— $d'$  accuracy versus processing time—for two conditions. Characteristically, the functions show a period of chance performance ( $d' = 0$ ), a period of increasing accuracy, and an asymptotic period during which further processing does not improve performance.

SAT asymptotes measure the probability (across trials and materials) that readers successfully arrived at an interpretation sufficient to support a positive response. Differences in asymptote, illustrated in Figure 1A, suggest that there are differences in the probability of successfully computing different interpretations. We predict lower asymptotes for coerced expressions if readers are more likely to fail to construct a sensible coerced interpretation or if they simply judge some of the resulting interpretations to be less acceptable.

The *intercept*, the point at which the SAT function departs from chance, and the *rate* at which the function grows to asymptote provide joint measures of processing speed. Differences in either parameter indicate that readers can interpret one expression more quickly than another (McElree, 1993; McElree & Nordlie, 1999). Figure 1B illustrates a pure rate difference. Whether speed differences are expressed in rate or intercept depends on the mean and variance of the difference in the time it takes to compute the respective interpretations. Importantly, in either case, the functions display disproportional dynamics, reaching a given proportion of their asymptote at different times. This provides a means of measuring differences in processing speed even when there are concomitant differences in accuracy. Thus, we can determine the relative speed of interpreting a coerced expression on the respective proportion of trials that readers succeed in computing a sensible interpretation.

The standard SAT procedure presents a single response (SR) signal at one of several times. Unfortunately, this requires many trials and sentences. The MR–SAT procedure provides a more efficient means of deriving SAT functions. Participants are trained to respond repeatedly within a trial to a series of (here, 14) tones at set (350-msec) intervals following the onset of the crucial expression and to dynamically modulate their responses to reflect changing assessments of the sentence. Findings of systematic dynamics differences with MR–SAT demonstrate the utility of the procedure (Bornkessel et al., 2004; McElree, 1993; Wickelgren et al., 1980). In addition, comparisons of the two procedures have shown that MR–SAT reproduces



**Figure 1.** Hypothetical speed–accuracy trade-off (SAT) functions illustrating two conditions that differ by (A) SAT asymptote only or (B) SAT rate. The intersection of the horizontal and vertical lines shows the point in time (abscissa) at which the functions reach two thirds of their respective asymptotes (ordinate). When dynamics are proportional (panel A), the functions reach the two-thirds point at the same time.

asymptotic and dynamics differences found in SR–SAT, as well as fine-grained shape differences between functions.<sup>1</sup>

### Two Types of Coerced Expressions

One set of materials contrasted sentences such as (a) and (b) in Table 1, analogous to some of our reading time studies. Sentences were presented phrase by phrase, and participants judged the sensibility of the sentences at times after the onset of the final phrase. At issue was whether the participants would process the final NP (e.g., *the table*) more quickly and more accurately following the control (e.g., *built*) than following the coercing (e.g., *began*) verb. Half the sentences were acceptable (a–c), and half were unacceptable (d–f) at the final phrase.

Another set examined coercion in a potentially more difficult construction. The main contrasts were (g) versus (h) and (i) versus (j) in Table 1. These constructions were formed around verbs that select for small clause complements, consisting of an NP subject and an adjectival predicate. The adjectives (e.g., *survivable*) were morphologically derived from event-selecting verbs—hence,

semantically required eventive subject NPs. Although not highly frequent forms, they follow standard morphological principles. We anticipated that morphologically complex adjectival phrases (APs) would be more difficult to process than the VPs in Sentences a–c, leading to lower accuracy and decreased processing speed. Nevertheless, composition should be simpler when a subject NP denotes an event (e.g., *the fall*) than when it does not (e.g., *the ice*). In the latter, the adjective should force coercion of the NP into an eventive interpretation (e.g., *climbing of the ice*). The APs generalize the phenomena to a different level of compositional difficulty and to a different construction. Specifically, coercion in the first set acts on a verb complement, whereas here it operates on the subject NP of an AP. For generality, we used two matrix verbs: those that can take a simple transitive form (e.g., *imagine*), and those that require a clausal complement (e.g., *prove*). Unacceptable versions (k and l) used an anomalous, noncoercible NP.

The V–NP set constituted 22% of the trials, as did the NP–AP set. Filler sentences of various types and lengths,

**Table 1**  
**Example Sentences**

Condition	Sentences
Acceptable sentences	V–NP
a. Coercing verb	The carpenter / began / the table.
b. Control verb	The carpenter / built / the table.
c. Additional filler	The babysitter / comforted / the child.
Unacceptable sentences	
d. Matched to NP	The carpenter / comforted / the table.
e. Matched to coercing verb	The burglar / began / the moon.
f. Matched to control verb	The hunter / built / the monsoon.
Acceptable sentences	NP–AP
g. Transitive verb / coerced NP	The climber / imagined / the ice survivable.
h. Transitive verb / control NP	The climber / imagined / the fall survivable.
i. S-clause verb / coerced NP	The climber / proved / the ice survivable.
j. S-clause verb / control NP	The climber / proved / the fall survivable.
Unacceptable sentences	
k. Transitive verb / unacceptable NP	The climber / imagined / the acuity survivable.
l. S-clause verb / unacceptable NP	The climber / proved / the acuity survivable.

Example filler types (asterisks = unacceptable versions; percentage = percentage of total set)

- The hair stylist / blow-dried / the straight hair. (5%)
- \*The hair stylist / blow-dried / the straight vision. (5%)
- The corrupt cop / spotted / the suspect loose. (5%)
- \*The corrupt cop / set / the pancake loose. (5%)
- Before jumping / into a taxi, / the reporter / and the candidate / had / a quick interview. (5%)
- \*Before jumping / into a taxi, / the reporter / and the candidate / had / a quick greenhouse. (5%)
- What / the older foreman / examined was / the unrolled blueprint. / It curled. (6.5%)
- \*What / the older foreman / examined was / the unrolled blueprint. / He curled. (6.5%)
- It was / the friendly barmaid / who served / the fancy cocktail. / She chattered. (6.5%)
- \*It was / the friendly barmaid / who served / the fancy cocktail. / I chattered. (6.5%)

Note—NP, noun phrase; AP, adjective phrase; S-clause verb, verb compatible with a sentential clause. Slashes denote the phrase breaks in the phrase-by-phrase presentation method.

including two-sentence structures, made up the remaining 55% (see Table 1). In total, 7.2% of the materials required coercion, making it unlikely that the participants would differentially focus on the coercion sentences.

## METHOD

### Participants

Twenty-six native-English speakers from New York University (NYU) participated in the study. They were paid \$40 for four 50-min sessions.

### Stimuli

We constructed 45, six-sentence contrasts for each of the V–NP and NP–AP materials. The V–NPs required a control verb expressing the eventive interpretation readers most often ascribed to the coerced sentences. An additional 24 NYU participants provided one- to two-word fill-in-the-blank responses indicating how they would interpret coerced sentences such as *The carpenter began \_\_\_ the table*. From 90 candidate sentences, we selected 45, using the dominant response (occurring, on average, 76% of the time) as the control verb (e.g., *built*). Francis and Kučera (1982) frequency counts averaged 218.2 occurrences per million for the coercing verbs and 332.4 for the control verbs. This difference was significant [ $t(44) = 2.21, p = .032$ ]; however, the coercing verbs occurred twice as often as the control verbs, which was likely to offset any frequency advantage. For diversity, 20 of the 45 V–NP clauses appeared as clausal complements of S-clause verbs (e.g., *The investigator believed that . . .*). Unacceptable variants of the sentences (a–c in Table 1) were created by generating anomalous NPs (e.g., d–f) or by using an inappropriate constituent in this position (e.g., . . . *began the clear*).

The NP–APs varied the subject NPs and, hence, could not be matched for meaning. Two experimenters generated what they regarded as plausible entity- and event-denoting subject NPs for the constructions. Frequency counts averaged 114.5 for the entity-denoting nouns and 87.2 for the event-denoting nouns [ $t(44) = 1.07, p = .29$ ].

If the control and coerced sentences varied in plausibility, we would expect this to be reflected in SAT asymptotes (lower plausibility engenders a lower proportion of positive sensicality judgments). Hence, SAT provides an indirect plausibility measure and obviates matching on plausibility ratings.<sup>2</sup> The 45 sets of V–NP and NP–AP materials were divided into 3 experimental sets, so that an equal number of the six conditions (a–f and g–l in Table 1) occurred in each set, but so that only 2 sentences in each sextuplet occurred in each set. The experimental sets were combined with 3 sets of 224 filler sentences.

### Procedure

A PC controlled stimulus presentation, timing, and response collection. A trial began with the 1-sec presentation of a random answer key symbol (1 = “unacceptable” or 3 = “acceptable”), which indicated which key the participants should start with. Phrases were presented one after another for durations equal to 300 msec/word. At the onset of the final phrase, a series of 14 auditory response cues (50-msec–1000-Hz tone) occurred every 350 msec.

The participants gave their current judgment of the sensicality of the sentence by pressing one of the two response keys, modulating their responses as their assessment changed. The participants had a 1-h practice session, which included training on dynamically modulating their responses. The experimental phase consisted of three 50-min sessions.

### Data Analysis

We calculated  $d'$  scores to correct for response bias, scaling  $z[P(\text{“yes”}|\text{acceptable})]$  against  $z[P(\text{“yes”}|\text{unacceptable})]$ . The hit rates from Conditions a and b in Table 1 were scaled against a common false alarm rate from the average of d, e, and f, and hit rates from Conditions g–j were scaled against the average of k and l. We calculated  $d'$ s in twelve 350-msec time bins surrounding the response cues (collapsing the first and last two bins).

The  $d'$  functions were fit with an exponential approach to a limit:

$$d'(t) = \lambda(1 - e^{-\beta(t-\delta)}), \text{ for } t > \delta, \text{ else } 0. \quad (1)$$

Here,  $\lambda$  indexes the function's asymptote,  $\delta$  the intercept, and  $\beta$  the rate of growth to asymptote. We fit the data with hierarchically nested models, ranging from a null model (all conditions were fit with a single  $\lambda$ ,  $\beta$ , and  $\delta$ ) to a fully saturated model (unique parameters

for each condition), using an iterative hill-climbing least-squares algorithm (Reed, 1976). Fit quality was assessed by an adjusted- $R^2$  statistic—the proportion of variance accounted for, adjusted by the number of free parameters (Judd & McClelland, 1989)—and by an evaluation of the consistency of the parameter estimates across the participants. To verify that differences were reliable across materials, we computed separate functions for each sentence (averaged across participants) and then fit each of the functions with Equation 1.

## RESULTS

### V–NP Sentences

Average accuracy across the three longest response bins provided an empirical measure of asymptotic performance. The asymptotes were higher for control (e.g., . . . *built the table*) than for coerced (e.g., . . . *began the table*) sentences by 0.35  $d'$  units, on average [ $t_1(25) = 4.07, p = .0003$ ;  $t_2(44) = 3.52, p = .001$ ]. Correspondingly, adequate fits of the time course functions required two asymptotic ( $\lambda$ ) parameters, and indeed fits with a single  $\lambda$  left systematic residuals at late processing times and yielded lower adjusted- $R^2$  values than did a  $2\lambda-1\beta-1\delta$  model. Crucially, differences in the estimated  $\lambda$  parameters were significant across participants [ $t_1(25) = 2.14, p = .04$ ] and items [ $t_2(44) = 2.29, p = .027$ ]. This asymptotic difference indicates that the readers were less likely to compute an acceptable interpretation of the coerced sentences.

Notably, the readers were also slower to interpret coerced expressions. There were no detectable differences in

intercept, but there were systematic rate ( $\beta$ ) differences. Although adjusted- $R^2$  values were only slightly higher for a  $2\lambda-2\beta-1\delta$  than for a  $2\lambda-1\beta-1\delta$  model, there were reliable differences in the estimated  $\beta$  parameters by participants [ $t_1(25) = 2.56, p = .017$ ] and by items [ $t_2(44) = 2.41, p = .02$ ]. In the average data,  $\beta$  was 1.15 for the coerced sentences and 1.34 for the control strings, a difference of 123 (1/ $\beta$ ) msec. Overall, a  $2\lambda-2\beta-1\delta$  model provided the best description of the time course profiles (adjusted- $R^2 = .998$ ; .891–.988 across participants), and it is plotted in Figure 2 for the average data.

### NP–AP Sentences

We excluded 2 participants from further analysis because their performance for NP–AP constructions was near the chance level. Even excluding these data, asymptotes for NP–APs were lower overall than those for V–NPs (2.66 vs. 3.27 in  $d'$  units). Hence, these constructions were indeed more difficult to interpret than V–NP constructions.

Nonetheless, a remarkably similar coercion effect was found. Control sentences (e.g., . . . *the fall survivable*) had higher asymptotes than did coerced sentences (e.g., . . . *the ice survivable*) by 0.31  $d'$  units on average. An effect of coercion was yielded by  $2 \times 2$  ANOVAs with verb type and coercion as within-participants and -items factors [ $F_1(1,23) = 18.15, p = .0003$ ;  $F_2(1,44) = 7.44, p = .009$ ]. The effect of verb—whether the matrix

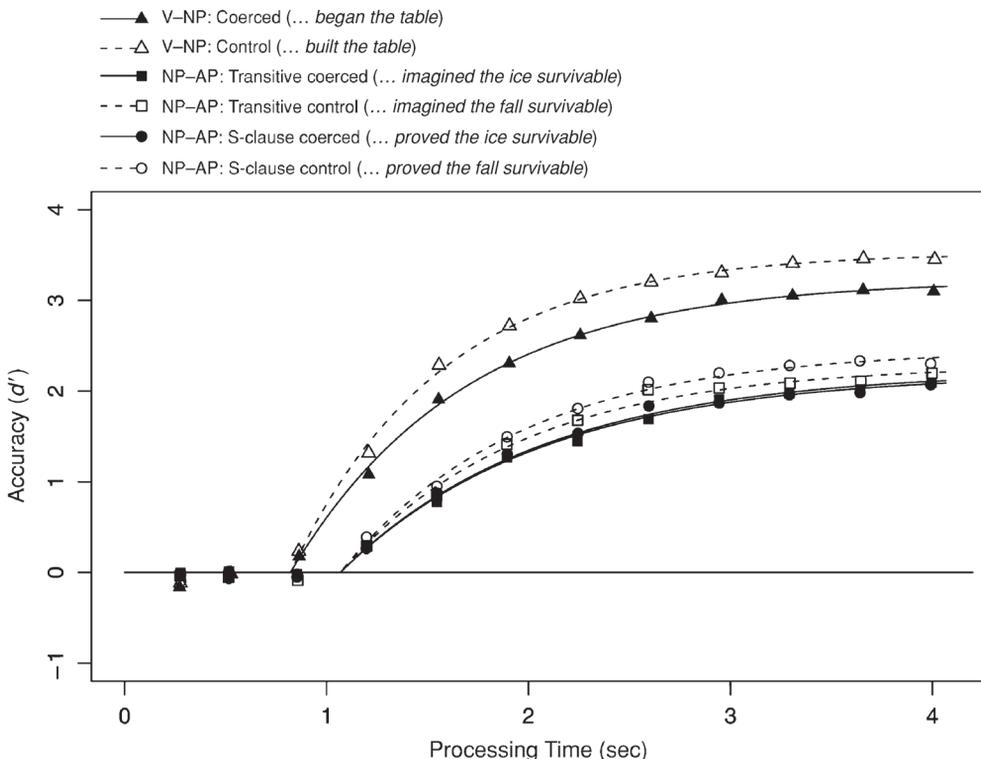


Figure 2. Average  $d'$  accuracy (symbols) as a function of processing time (350-msec time bins following the final phrase) for judgments of verb–noun-phrase (V–NP; triangles) and NP–adjective-phrase (NP–AP; squares and circles) sentences. Smooth curves show the best-fitting exponential models (see the text).

verb strictly required a clausal complement—was not significant ( $F_s \approx 1$ ). There were hints of an interaction [ $F_1(1,23) = 5.43, p = .03; F_2(1,44) = 1.66, p = .20$ ], since asymptotes were lower for verbs that optionally, as compared with strictly, required complement clauses in the control sentences (2.23 vs. 2.39  $d'$  units), but not in coerced sentences (2.11 vs. 2.08  $d'$  units). This could reflect a slight tendency to misparse sentences with verbs with optional subcategorizations. The  $\lambda$  estimates from model fits reflected the empirical pattern: When asymptotes were varied with condition, there was a significant coercion effect in the  $\lambda$  estimates [ $F_1(1,23) = 18.7, p = .0003; F_2(1,44) = 12.5, p = .001$ ], no effect of verb type ( $F_s \approx 1$ ), and questionable evidence for an interaction [ $F_1(1,23) = 3.24, p = .08; F_2(1,44) = 0.707, p = .40$ ]. Importantly, the lower asymptotes for coerced sentences indicate that the readers were less likely to compute a sensible enriched interpretation.

Here, too, coercion affected processing speed, since the participants had slower rates for coerced than for control sentences. (Dynamics did not vary with the matrix verb type, either as a main effect or in interaction with coercion.) In the  $4\lambda - 2\beta - 1\delta$  model, the  $\beta$  estimates were significantly lower for coerced than for control sentences [ $t_1(23) = 3.16, p = .04; t_2(44) = 2.17, p = .035$ ]. For the average data,  $\beta$  was 0.91 for the coerced sentences, as compared with 1.06 for the controls, a difference of 158 ( $1/\beta$ ) msec. The  $4\lambda - 2\beta - 1\delta$  model provided the best description of the time course profiles (adjusted- $R^2 = .997$ , ranging from .733 to .981) and is shown in Figure 2.

## DISCUSSION

Expressions thought to require coercion yielded lower asymptotic levels than did minimally contrasting controls, whether coercion was triggered by an event-selecting verb (V–NP) or an adjective (NP–AP). These asymptotic differences suggest that the readers were less likely to compute sensible eventive interpretations of the NPs. This might have been the case if, on some trials, the readers did not attempt to coerce the subject or complement NPs, attempted but failed to derive any eventive interpretation, or derived what they judged to be an unacceptable interpretation.

That the readers were less likely to compute an acceptable interpretation for a coerced expression could simply indicate that coerced interpretations are typically less constrained than were the control interpretations. In this respect, one might view this difference as similar to the difference between metaphoric and literal expressions, where readers are less likely to recover or compose the information that is needed to construct a metaphoric interpretation (McElree & Nordlie, 1999). However, unlike metaphors, coerced expressions were also processed more slowly than their controls, with average differences of 123 msec in V–NP constructions and 158 msec in NP–AP constructions. These dynamics differences go beyond reading time findings by providing direct evidence that readers require more time to construct a sensible coerced interpretation.

One interpretation of this speed difference is that readers first deploy standard compositional operations in an attempt to unify the semantic properties of the NP with the verb or adjective but, failing to derive an acceptable interpretation, engage additional processes to construct an alternative sense of the NP. However, serial models typically predict larger dynamics effects than those observed here and often engender intercept differences (for simulations, see McElree, 1993; McElree & Carrasco, 1999; McElree & Doshier, 1993). The time course profiles are more compatible with the idea that mismatching properties between the NP and the verb or adjective immediately trigger more time-consuming coercion operations. Overall, the findings are consistent with our account of reading time effects: Coercion operations require more time because readers must engage processes to build additional semantic structure to instantiate an event sense for the NP. This account is also consistent with the finding that the time course differences—123 msec for V–NP coercions and 158 msec for NP–AP coercions—were remarkably similar across the different constructions, despite sizable differences in accuracy (0.6  $d'$  units) and in processing speed (average processing speed,  $\delta + 1/\beta$ , was 1,629 msec for V–NPs and 2,106 msec for NP–APs). This pattern is consistent with the deployment of a common coercion operation. A unified account of the full time course pattern follows from a natural assumption that these more time-intensive, complex compositional operations (coercion) also increase the likelihood of an error—hence, lower asymptotic performance.

## CONCLUSION

The SAT results show that enriched composition has measurable effects on the speed of interpretation above and beyond differences in the likelihood of constructing a sensible interpretation. Collectively, the time course measures and other reading time results indicate that composition can entail more than combining retrieved semantic properties. Often, it involves operations that generate unstated semantic content, which extends the sense of an expression so that a reader can establish an internally coherent and plausible interpretation.

## REFERENCES

- BORNKESSEL, I., McELREE, B., SCHLESEWSKY, M., & FRIEDERICI, A. D. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from case marking. *Journal of Memory & Language*, *51*, 495-522.
- FRANCIS, W. N., & KUČERA, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- JACKENDOFF, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.
- JUDD, C. M., & McCLELLAND, G. H. (1989). *Data analysis: A model-comparison approach*. San Diego: Harcourt Brace Jovanovich.
- LASCARIDES, A., & COPESTAKE, A. (1998). Pragmatics and word meaning. *Journal of Linguistics*, *34*, 387-414.
- McELREE, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory & Language*, *32*, 536-571.
- McELREE, B., & CARRASCO, M. (1999). The temporal dynamics of vi-

- sual search: Evidence for parallel processing in feature and conjunction searches. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 1517-1539.
- McELREE, B., & DOSHER, B. A. (1989). Serial position and set size in short-term memory: Time course of recognition. *Journal of Experimental Psychology: General*, **18**, 346-373.
- McELREE, B., & DOSHER, B. A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General*, **122**, 291-315.
- McELREE, B., & NORDLIE, J. (1999). Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review*, **6**, 486-494.
- McELREE, B., TRAXLER, M. J., PICKERING, M. J., SEELY, R. E., & JACKENDOFF, R. (2001). Reading time evidence for enriched composition. *Cognition*, **78**, B17-B25.
- PICKERING, M. J., McELREE, B., & TRAXLER, M. J. (2005). The difficulty of coercion: A response to de Almeida. *Brain & Language*, **93**, 1-9.
- PUSTEJOVSKY, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
- REED, A. V. (1976). List length and the time course of recognition in immediate memory. *Memory & Cognition*, **4**, 16-30.
- TRAXLER, M. J., McELREE, B., WILLIAMS, R. S., & PICKERING, M. J. (2005). Context effects in coercion: Evidence from eye movements. *Journal of Memory & Language*, **53**, 1-25.
- TRAXLER, M. J., PICKERING, M. J., & McELREE, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory & Language*, **47**, 530-547.
- WICKELGREN, W. A., CORBETT, A. T., & DOSHER, B. A. (1980). Priming and retrieval from short-term memory: A speed-accuracy tradeoff analysis. *Journal of Verbal Learning & Verbal Behavior*, **19**, 387-404.

#### NOTES

1. McElree and Doshier (1989) replicated in SR-SAT the asymptotic and dynamics patterns in the MR-SAT study of Wickelgren et al. (1980). In addition, within-subjects comparisons showed that MR-SAT reproduces SR-SAT asymptotic and dynamics differences, even though absolute speed estimates were 25% slower in MR-SAT. The slower dynamics in MR-SAT is consistent with the increased response load and with modest effects of hysteresis.

2. Nonetheless, another 20 NYU students rated the plausibility of the V-NP coerced (a) and control (b) sentences on a 7-point scale (7 = *highly plausible*). Coerced sentences averaged 6.25 and control sentences 6.58 [ $t(44) = 1.4, p = .17$ ].

(Manuscript received August 3, 2004;  
revision accepted for publication May 5, 2005.)