Developmental Science 7:3 (2004), pp 360–377

PAPER

The development of complex sentence interpretation in typically developing children compared with children with specific language impairments or early unilateral focal lesions

Frederic Dick,[1,2] Beverly Wulfeck,[1,4] Magda Krupa-Kwiatkowski[1] and Elizabeth Bates[1,3]

1. *Center for Research in Language, University of California San Diego, USA*
2. *Department of Psychology, Birkbeck College, University of London, UK*
3. *Department of Cognitive Science, University of California, USA*
4. *School of Speech, Language, and Hearing Sciences, San Diego State University, USA*

Abstract

This study compared sentence comprehension skills in typically developing children 5–17 years of age, children with language impairment (LI) and children with focal brain injuries (FL) acquired in the prelperinatal period. Participants were asked to process sentences 'on-line', choosing the agent in sentences that varied in syntactic complexity (actives, passives, subject clefts and object clefts), and in the presence or absence of a subject-verb agreement contrast. Results revealed that accuracy and processing speeds vary with syntactic complexity in all groups, reflecting the frequency and regularity of sentence types. Developmental changes continued throughout childhood, as children became faster and more accurate at processing more complex sentence structures. Children with LI and children with FL were quite profoundly delayed, displaying profiles similar to, or more impaired than those of younger children, but there was no evidence in the FL group for a disadvantage in left- vs. right-hemisphere-damaged children. Children with LI showed one unique pattern: higher than normal costs (reflected in reaction times) in using converging information from subject-verb agreement, in line with studies suggesting special vulnerabilities in grammatical morphology in this group. Results are discussed in terms of the Competition Model, a theory of language processing designed to account for the statistical changes in performance that are observed during development, and the probabilistic deficits in children with language impairments.

Introduction

Language development is not restricted to the acquisition of words and/or rules. To comprehend language, a child must engage in very rapid processing of phonological, lexical/semantic, grammatical and syntactic information presented by the speaker. To be successful, the child must take advantage of context in order to access and integrate information over multiple levels, with millisecond timing. Given these constraints, it is clear that language development involves much more than the acquisition of knowledge. Children must also develop the ability to activate and deploy that knowledge efficiently in real time. The development of language processing efficiency takes place across the school years and beyond, and requires refinement of the temporal calibration of the language as well as the reanalysis of language knowledge for discourse purposes. In the same vein, language impairment in children within the same age range is likely to involve more than the failure to acquire knowledge. Even though a child eventually may 'know' the grammatical structures of his/her language, s/he may fail to use that knowledge efficiently in real time.

Although there is an extensive literature on early language acquisition (e.g. Bloom & Lahey, 1978; Bloom, Lahey, Hood, Lifter & Fiess, 1980; Brown, 1973; Fletcher & MacWhinney, 1995; Hayes, 1970; Owens, 1991; Reich, 1986; Wiig & Semel, 1984), far less is known about language in later stages of development, including the kind of 'fine-tuning' of the skill system that must be done for

Address for correspondence: Frederic Dick, Brain Research Imaging Center, Department of Neurology, University of Chicago, MC-2030, 5841 S. Maryland Av., Chicago, IL 60637, USA; e-mail: fdick@crl.ucsd.edu

efficient language use. Until recently, most studies of language comprehension in children have used off-line paradigms that do not tap the real-time properties of language processing, but instead focus on post-sentence performance. However, in most communicative interactions, people rarely have time for reflection as they communicate. If we are to understand the complex mechanisms which underlie language representation and processing, then we need to study language *during* as well as *after* processing.

The purpose of the present study is to investigate normal and atypical development of the ability to process complex sentences. The study employs an on-line sentence interpretation paradigm to investigate the temporal microstructure of auditory language processing in order to document developmental changes in real-time sentence processing. We focus on the period from 5 years to young adulthood, in typically developing children and in two groups at risk for protracted language delays: children with behaviorally defined language impairment, and children with congenital injuries to one hemisphere of the brain. The proposed study is an extension of prior research examining language processing in children (Von Berger, Wulfeck, Bates & Fink, 1996), this time using complex sentence types that were not employed in the earlier study of typically developing children, and extending those methods to children at risk for, or with diagnosed language disorders. Before presenting the experimental details, we review the theoretical basis for this study as well as the 'on-line' sentence processing literature on typical and atypical populations of children.

Theoretical rationale: the Competition Model

This study is based on more than two decades of cross-linguistic and cross-age studies conducted by Bates, MacWhinney and colleagues, working within the framework of the Competition Model (CM), an interactive-activation model of sentence processing that has been developed to account for qualitative and quantitative differences in processing by child and adult speakers of different language types (Bates, Devescovi & Wulfeck, 2001; Bates & MacWhinney, 1987; MacWhinney, 1987; MacWhinney & Bates, 1989). An important aspect of the Competition Model is its emphasis on quantitative variation. Although some researchers have acknowledged the role of processing factors (Bloom, Miller & Hood, 1975) and variation (Cedergren & Sankoff, 1974; Labov, 1969) in language development, most theories of language have traditionally focused on linguistic competence. Language acquisition is described in terms of rules or regularities that are either present or absent, acquired by children in discrete steps and applied by adults in an all-or-nothing fashion at various points across the course of a sentence (c.f. Fodor, Bever & Garrett, 1974). In such models, statistical variation is treated as noise or as the product of some unspecified performance factor.

By contrast, the Competition Model is designed to account for the probabilistic variations that characterize linguistic performance by adults, and for the statistical changes in performance that are observed over time in children. It is an early version of the kinds of connectionist or constraint-satisfaction models that are now quite common within adult psycholinguistics, models that emphasize distributed and probabilistic links within the microstructure of language (Elman, 1990; Kempe & MacWhinney, 1999; MacDonald, 1999; Manning & Schütze, 1999). As is the case in all connectionist models, linguistic knowledge in the Competition Model is characterized not as a set of rules but as a complex network of weighted form-meaning mappings. Every language must provide cues (lexical, syntactic, morphological or prosodic) that signal the presence of universal meanings (e.g. the agent role). These mappings can vary in strength within and across languages, e.g. word order is an extraordinarily strong cue to the agent role in English, whereas subject-verb agreement is the most important cue in Italian. In contrast with competence-based models (which predict yes-or-no mappings), the Competition Model allows for a full range of probabilistic values between these two extremes (e.g. the noun before a verb is usually, though not always, the agent of an action). Many different sources of information are examined and weighed together across the course of the sentence, until the system settles into the best available fit between meaning and form. From this point of view, language development involves more than acquiring rules or mappings; it also involves a gradual process of the tuning of these mappings to fit the profile of strengths and weaknesses reflected in the linguistic input. Indeed, data showing a sharply discontinuous change in a child's ability to use a particular grammatical structure (or conversely a complete and selective inability with that structure) would run opposite to the model's predictions.

The major predictive construct in the Competition Model is *cue validity*, which refers to the information value of a given source of information (e.g. preverbal position) for a particular communicative function or meaning (e.g. the agent role). Cue validity has been analyzed into two components: *cue availability* (how often is a particular cue available when we need it to assign an agent role?) and *cue reliability* (when the cue is available, how often does it lead to the right answer?). All other things being equal, the order in which form-function mappings are acquired will reflect the relative strength of that mapping, with the most valid cues acquired first. When these mappings are in place, children adjust them until they provide an optimal fit to the processing environment (i.e. cue strengths reflect cue validity). In other words, cue strength is a characteristic of the processing device (in this case, the child); with learning, cue strength is increasingly correlated with cue validity, a property of the linguistic environment.

It is important to note that cue validity is related to frequency, but it is not the same thing. Three specific characteristics of cue validity are especially important for the study to be presented here.

First, cue validity is defined as the ratio of reliability over availability. That means that the denominator is driven by function rather than raw frequency. In the present study, we will concentrate on syntactic and morphological cues to the agent role in English, where validity is a function of how often agent-role decisions have to be made.

Second, cue validity is mitigated by a second major factor called *cue cost*, which refers to the processing costs involved in the use of any given cue. Two cues can be equivalent in their overall information value or validity (e.g. two accusative case marking cues in Hungarian), but differ in their perceivability, retrievability, pronounceability, memorability, and so forth (e.g. one form of the accusative marker in Hungarian ends in a consonant cluster, another ends in a consonant preceded by a strong vowel, which makes the second form easier to hear). Under those conditions, cue cost may create a difference both in the degree to which adult listeners rely on the cue in comprehension, and in the timing of language development for the same structures (as has been shown to be true for the 'easy' vs. 'hard' accusative case markers in Hungarian – MacWhinney, Osman-Sagi & Slobin, 1991; MacWhinney & Pleh, 1988). In the present study, we will cross contrasts in word order types (actives, passives, subject clefts and object clefts) with contrasts in the availability of subject-verb agreement cues. As reviewed by Bates, Wulfeck and MacWhinney (1991), subject-verb agreement cues are not only relatively low in validity in English (they are much less reliable than word order as cues to the agent role), they are also low in perceptual salience, which results in high processing costs. Hence we may expect agreement cues to be especially vulnerable during development in children at risk for language disorders.

Third and finally, because the Competition Model assumes distributed representations, both the notion of 'agent' and the cues that signal that meaning are viewed not as unitary phenomena, but as coalitions of finer-grained patterns of sound and meaning. In other words, cues and their meanings have 'microstructure' (Rumelhart & McClelland, 1986) – and both adult processing and acquisition in children are presumed to be driven by cue validity and cue cost in this microstructure. This means that a macrostructure like the subject cleft (e.g. 'It's the dog that is kicking the cow'), which is in itself low in overall frequency, may be high in validity because it overlaps extensively in structure with the high frequency active (Subject-Verb-Object) sentence types (e.g. 'The dog is kicking the cow'). In contrast, the object cleft (e.g. 'It's the dog that the cow is kicking') is not only low in absolute frequency, but contains substructures (Object-Subject-Verb) that are also low in frequency and validity in English. The same thing is true for the passive (e.g. 'The dog is kicked by the cow'), with its lower-frequency and lower-validity Object-Verb-Subject structure. However, the passive contains some salient and highly valid substructure (e.g. the verb participial form 'is kicked' and the preposition by-phrase 'by the cow') which may make the passive relatively easier than the object cleft form, for both adults and children.

On-line studies of sentence comprehension in typical and atypical children

The original sentence comprehension procedure, upon which the present and many previous sentence interpretation studies have been based, can be thought of as a 'whodunit task', in that the research subject serves as a detective whose job it is to uncover the agent of the action. For example, in *off-line* versions, subjects are asked to interpret simple sentences by acting them out with small toy objects (e.g. Show me 'The cow is kicking the pencil'). The sentence stimuli represent converging and competing combinations of semantic cues (i.e. the contrast between animate and inanimate objects), syntactic cues (i.e. canonical and non-canonical word orders) and morphological cues (i.e. presence/absence of subject-verb agreement with the first or second noun). For example, in English noun-verb-noun sentences, the first noun is more likely to be the agent of the sentence than the second noun, and animate objects are more likely to be agents compared to inanimate objects. This design permits assessment of the hierarchy of importance of syntactic, semantic and morphological cues to agent/object relations, and serves as a direct test of the cue validity predictions of the Competition Model.

In one of the first studies using this paradigm (Bates *et al.*, 1984), it was shown that English-speaking children rely on word order by 2 years of age, a tendency that increases markedly by age 5. In particular, the results revealed early emergence of canonical Subject-Verb-Object (SVO) interpretations with noun-verb-noun constructions. In contrast, there was a marked delay in the emergence of non-canonical Object-Subject-Verb (OSV) and Verb-Object-Subject (VOS) interpretation strategies. The latter strategies are quite robust for English-speaking adults in the interpretation of noun-noun-verb and verb-noun-noun constructions, respectively.

The Bates *et al.* results are compatible with several earlier studies demonstrating the protracted development of non-canonical sentence forms across childhood. For example, in a sentence interpretation study of young children, Bever (1970) found clear evidence that children used an SVO strategy interpreting noun-verb-noun (NVN) surface structures. He proposed that this might be due to the fact that these types of structures occur more frequently in the language. Bever (1970) also observed that children tended to over-generalize the SVO strategy with more complex structures such as passive, reversible passive and reversible object cleft sentences. Bever (1970) suggested that through experience with these more complex structures older children learn to process them accurately. Turner and Rommetveit (1968) also found that more complex forms tend to be learned after the child begins school as opposed to learning them from everyday speech. In this study, the researchers investigated the comprehension abilities for non-reversible and reversible active and passive sentences in grade school children. They found an order of difficulty for these structures. Best performance was noted on non-reversible actives, then reversible actives, non-reversible passives, and finally reversible passives. They also found that as children advanced in age they were not only more accurate in comprehending more complex structures like reversible passives, but were also more consistent in their responses.

Slobin (1966) was among the first to examine *on-line* sentence comprehension of semantically reversible and non-reversible active, passive and negative forms in children between 6 and 12 years of age. He noted the following: (1) errors decreased with age; (2) longer response times (RT) were associated with incorrect responses; (3) passives and negative forms presented the most difficulty for all children; (4) faster RTs were obtained on non-reversible sentences compared to reversible sentences; and (5) decreased RTs were observed across age; however, the smallest differences between sentence types were noted for the oldest children (ages 10–12 years). Slobin concluded that, throughout childhood, there is continuous development of language learning (see also Tyler & Marslen-Wilson, 1981).

In a study of the contributions of verbal and spatial working memory to children's (ages 5–8) sentence comprehension, Roe (2003) not only showed that younger children were particularly inaccurate in comprehending more complex syntactic structures (e.g. a complexity × age interaction), but that children with smaller verbal working memory spans were slower and less accurate in comprehending these more complex structures than were their higher-span counterparts (e.g. a memory span × complexity interaction). Indeed, verbal working memory accounted for much more of the variance in reaction time and accuracy measures than did chronological age. In a similar vein, Booth, MacWhinney and Harasaki (2000) showed that high- and low-digit-span children (ages 8–11) employed different processing (attachment) strategies when comprehending sentences with relative clauses.

Wulfeck (1993) investigated grammaticality judgments and decision times for two age groups of typically developing children to determine when parsing decisions are made and how linguistic knowledge affects parsing. Results revealed that children showed good sensitivity to grammatical violations but not the ceiling performance observed in adults. Moreover, while adults showed no differential sensitivity to violation type, children were better at detecting violations created by permuting words (e.g. 'At the church by the park people standing are') in a sentence compared to their ability to recognize errors of morphological selection (e.g. 'The teacher have taken a small shell from the bucket'). This suggests that developmental changes in sensitivity to different aspects of morphology take place during the elementary school years (e.g. word order, subject-verb agreement, number agreement). Although older children processed violations more quickly overall, both groups of children demonstrated very rapid integration of information during sentence processing. Greater sensitivity to word order violations seemed to enhance children's ability to take advantage of context across a sentence, resulting in faster decision times. In short, sensitivity and processing abilities were best for those structures (in this case, word order) that had the greatest cue validity in their language (in this case, English).

One of the first on-line studies of children within the Competition Model was conducted by Von Berger *et al.* (1996). They examined the syntactic, semantic and morphologic cues that 7–12-year-old children used to process sentences, measuring both agent choice and reaction times. These results were then compared to the cues used and the time required by adults, with a special focus on the emergence of non-canonical 'second-noun strategies' (i.e. OSV and VOS). The findings of their study were those predicted by the Competition Model. At all ages, subjects tended to prefer the first noun as the subject or agent of the Noun-Verb-Noun (NVN) structures. In other words, children interpreted these forms as SVO structures. None of the 7- to 8-year-olds exhibited any second noun strategies. The other age groups appeared to have some knowledge of these strategies, but use of these cues was not at the adult level. Children in the study also seemed to pay very little attention to other cues contained in the sentence (i.e. animacy or agreement). Finally, Von Berger *et al.* (1996) found that when cues within a sentence converged (e.g. *animacy* and *first noun*, as in 'The cow is kicking the pencil'), the response times decreased; if they competed (e.g. *first noun* and *non-animacy*, as in 'The pencil is kicking the cow'), response times increased. The findings of this study strongly corroborate the predictions of the Competition Model, and suggest that there are developmental changes in language processing well into the school years.

On-line or off-line developmental studies have now been conducted within the Competition Model in a large number of different languages (for reviews, see MacWhinney & Bates, 1989; Bates, Devescovi & Wulfeck, 2001; Devescovi, D'Amico & Gentile, 1999). These studies show above all that all of the effects we have just reviewed are peculiar to English. Other languages show different sequences and rates of development, with different cue combinations, including early or late usage of word order cues, and early or late usage of various morphological cues. In other words, the early emergence of SVO in English is not a universal phenomenon, nor is the late emergence of passives and alternative word order types (Demuth, 1989). Instead, order of emergence and strength of cues seems to depend primarily on the cue validity and cue cost factors reviewed above.

Turning to studies of atypical populations using the same designs, most studies suggest that impairment in on-line tasks is often best characterized in terms of delay rather than deviance – see papers in Wulfeck and Reilly (in press) and Leonard (1998). For example, Feldman, MacWhinney and Sacco (2002) used a design similar to that of Von Berger *et al.* (1996) to study on-line comprehension in 141 typically developing children from 5 to 12 years of age, and 15 children with congenital unilateral brain injuries (12 to the left hemisphere, three to the right). As a group, the focal lesion group fell behind typical children, showing later emergence of second noun strategies and more persistent reliance on animacy in both agent choice and reaction times – characteristics of processing generally associated with younger children. (Note that the same pattern was observed in two of three children with right hemisphere damage, as well as the left-hemisphere group, thus providing no evidence for early left-hemisphere specialization.)

Our group has conducted a series of studies comparing both language production and language comprehension in children with focal brain injury, children with language impairment of unknown origin, and typically developing controls (Wulfeck, Bates, Krupa-Kwiatkowski & Saltzman, 2003, Reilly, Bates & Marchman, 2003) as well as some comparisons involving children with Williams Syndrome (Weckerly, Wulfeck & Reilly, 2003). The most common finding, in study after study, confirms a pattern of delay rather than deviance in clinical populations. For instance, Weckerly *et al.* (2003) found that language-impaired (LI) children and children with either Williams Syndrome or early focal lesions (FL) showed the same profile and number of error types as younger typically developing (TD) children on a complex production task ('tag' questions). In an analysis of children's narratives (drawing from the same clinical groups and typically developing children), Reilly, Losh, Bellugi and Wulfeck (2003) showed that younger LI and FL children were indeed producing more morphosyntactic errors and less complex constructions than their age-matched TD counterparts, but that this difference was no longer in evidence at older (10–12) ages for FL children, suggesting a period of protracted but subsequently rapid development for these brain-damaged children. Despite the fact that LI children showed a more protracted delay, these children, as well as children with FL and Williams Syndrome, showed patterns of morphosyntactic errors that were similar to, and not deviant from, younger typically developing children – results that are predicted by the Competition Model, but not by more modular theories suggesting that language disorders can be best characterized by loss of specific grammatical knowledge or representation.

These and other studies tend to show that children with focal brain injury tend to perform below typically developing controls (in the low-normal range) but they also invariably outperform children with language impairment and/or children with Williams Syndrome. This kind of result provides strong evidence for behavioral and neural plasticity in the face of early brain injury, and suggests that the deficits underlying other forms of language impairment may be less plastic, perhaps reflecting a more diffuse form of brain injury. Note, however, that this hypothesis has not yet been tested against the kinds of complex sentences that we will be using here. When the stakes are raised, we may find that children with focal brain injury are still struggling just as much as children with LI to achieve adult-like levels of performance.

Finally, the present study will permit us to assess not only the comprehension of complex sentence types (object clefts, subject clefts, passives), but also the role of subject-verb agreement cues as aids to comprehension. A large literature on language impairment in children leads to the conclusion that grammatical morphology is a special area of vulnerability for this group. There is still a great deal of controversy regarding the causes of this special vulnerability: Is it due to genetically based damage to grammar-specific skills (van der Lely, Rosen

Table 1 *Children with focal lesions: age, gender and lesion location*

ID	Sex[a]	Age[b]	Hemi[c]	Lobe[d]	SC[e]	TH[f]	BG[g]	Med[h]
1	M	7	L	FTPO	X	X	X	X
2	F	7	L	SC	X		X	
3	F	7	L	FTPO	X	X	X	
4	F	8	L	FT	X			
5	M	10	L	FTP	X	X	X	
6	F	10	L	TPO	X	X	X	
7	M	11	L	T	X	X	X	
8	M	12	L	F	X			X
9	F	12	L	PO				
10	M	12	L	FTP	X			
11	M	18	L	FTPO	X	X	X	
12	M	7	R	P	X			
13	M	7	R	F				
14	M	7	R	TP				
15	M	9	R	FTPO	X	X	X	X
16	F	12	R	FTPO	X	X	X	
17	M	12	R	SC			X	
18	F	14	R	TP	X			
19	M	14	R	FTP	X	X	X	X
20	M	16	R	P	X	X		

[a] Abbreviations for sex are: M = Male, F = Female.
[b] Age is rounded to the nearest year.
[c] Abbreviations for hemisphere of injury: L = Left, R = Right.
[d] Abbreviations for lobe of injury: F = Frontal, T = Temporal, P = Parietal, O = Occipital, SC = Subcortical damage only.
[e–g] An 'X' indicates that the child has unilateral damage to Subcortical structures (SC), the Thalamus (TH), or Basal Ganglia (BG).
[h] An 'X' indicates that the child has taken or is currently taking psychoactive medications to control seizures or other neurological conditions.

& McClelland, 1998)? Or does it reflect the fact that grammatical morphemes constitute 'weak links in the processing chain', vulnerable not only to brain injury and/or genetic defects, but to deficits in information processing that can be simulated in normal adults forced to interpret linguistic material under various forms of perceptual degradation and/or cognitive overload (Dick, Bates, Wulfeck, Utman, Dronkers & Gernsbacher, 2001; Blackwell & Bates, 1995; Kilborn, 1991)? The present study will help to shed light on these issues.

Method

Participants

Participants were 102 typically developing children (TD – ages 5–17, mean 10.8 years), 24 language-impaired children (LI – ages 7–15, mean 9.3 years) and 20 children with early unilateral focal brain lesions (FL – ages 7–18, mean 10.8 years). The parents of the TD children completed questionnaires confirming normal developmental and educational histories and grade level performance in school. Children with language impairment (LI) had a documented language impairment and were recruited from speech-language pathologists and physicians. Language impairment was defined following established diagnostic criteria (Tager-Flusberg & Cooper, 1999): No evidence of frank neurological impairment (as determined by neurological exam), autism or social-emotional disturbance, a performance IQ score of 80 or better on the WISC-R or WISC-III, and an Expressive Language Composite Score-ELS (CELF-R or CELF-III) that was 1.5 or more standard deviations below the mean for that age group. Three of the 24 LI children had been or were being treated with psychoactive medications.

Children with early focal lesions (FL) were recruited from pediatric neurologists and pediatricians. All showed evidence of a *unilateral* left- or right-hemisphere focal lesion (as determined by CT or MRI – see Table 1 for patient information); lesion onset in all children was prenatal, perinatal or within the first 6 months of life (e.g. no children with traumatic head injury were included). All children had normal or corrected-to-normal vision and normal hearing. All participants were right-handed native English speakers and all were treated in accordance with the 'Ethical Principles of Psychologists and Code of Conduct' (American Psychological Association, 1992).

Design and materials

The sentence interpretation experiment was a 2 within-subjects factorial design where factors were Sentence Type, with four levels: (1) Active, (2) Subject Cleft, (3) Object Cleft and (4) Passive, and Subject-Verb Agreement Cue, with two levels: (1) Present and (2) Absent.

Table 2 *Example sentence types*

Sentence type (constituent order)	Agreement cue (Yes/No)	Example sentence (agreement cues underlined/bold)
Active (SVO)	No	The dog is biting the cat.
Active (SVO)	Yes	The dog**s are** biting the cat.
Subject Cleft (SVO)	No	It's the dogs that are biting the cats.
Subject Cleft (SVO)	Yes	It's the dog that **is** biting the cat**s**.
Object Cleft (OSV)	No	It's the cat that the dog is biting.
Object Cleft (OSV)	Yes	It's the cat that the dog**s are** biting.
Passive (OVS)	No	The cat is bitten by the dog.
Passive (OVS)	Yes	The cat **is** bitten by the dog**s**.

Between-subject variables were Age in Years (5–6, 7–8, 9–10, 11–12, 13–14, 15-up) and Population (TD, FL, LI); for some analyses involving children with focal lesions, Hemisphere Damaged (Left, Right) is also included as a between-subjects variable.

Dependent variables were Percent Correct Response (%CR) and Mean Reaction Time (RT). Reaction times were corrected for two potential confounds, sentence length and motor response latency. The length of the sentence stimuli varied systematically across type, where mean sentence lengths in milliseconds were: Actives = 2117, Subject Cleft = 2367, Object Clefts = 2387, Passives = 2184. Gross motor response latencies (as measured by a baseline reaction time task – see below) also tended to vary widely across ages and subject groups, thereby potentially skewing results. Therefore, for every datapoint, we subtracted the specific sentence length as well as the subject's mean baseline response time before computing summary statistics. In addition, we included reaction times only from correct responses in mean RT calculations.

Experimental materials consisted of both visual and auditory stimuli. Visual stimuli were 3″ × 2″ digitized black-and-white line drawings of familiar animals culled from several picture databases (Abbate & LaChapelle, 1984a, 1984b; Snodgrass & Vanderwart, 1980). Displayed on a VGA color monitor, each drawing was embedded in a solid gray rectangle over a white background; drawings were presented in pairs determined by sentence content, and projected to the left and right sides of the monitor.

Sentence stimuli consisted of 96 sentences that were generated by first randomly assigning two animate nouns (from a pool of 12) to one transitive verb (from a pool of 15). All 12 nouns referred to familiar animals, and all could be assigned to either agent or patient roles. All 15 verbs were semantically similar, in that they expressed a 'bad action', such as chasing or hurting. Twenty-four noun-verb pairs were then randomly assigned to each of the four syntactic structures: Active, Subject Cleft, Object Cleft and Passive. Each of these syntactically marked pairs was then pseudorandomly assigned to one of four inflectional paradigms: (1) subject and object inflected in singular, verb agrees with both; (2) subject singular, object plural, verb agrees with subject; (3) subject plural, object singular, verb agrees with subject; (4) both subject and object plural, verb agrees with both. Each level of the Sentence Type variable was thereby represented by 24 exemplars, half of which contained a cue to agency via subject-verb agreement (inflections (2) and (3)), and half of which contained no agreement cue to agency (inflections (1) and (4)) – see Table 2 for example sentences. The present progressive form of the verb was used for all 96 sentences to retain continuity with related studies (e.g. Dick *et al.*, 2001). The complete stimulus set can be viewed online at http://crl.ucsd.edu/~fdick/devcsi/sents.html, including sentence durations.

Sentence stimuli were digitally recorded in a sound-insulated chamber by an experienced female speaker (EB), and were read with a smooth and neutral intonation across the different sentence types. Two independent raters listened to each sentence and those sentences that were judged as having biasing or contrastive stress were re-recorded. Recordings were then converted to SoundEdit16 files, with a 22.255 kHz sampling rate and 8-bit quantization.

Equipment

PsyScope software (version 1.0.1 and version 1.0.2) was used to deliver stimuli and collect data (Cohen, MacWhinney, Flatt & Provost, 1993). Software was run on Macintosh Performa 6214 computers connected to a VGA color monitor and Apple external speakers (AppleDesign Powered, with tuned port bass reflex speakers). A PsyScope button box was used for response and experimental timing.

Procedure

Participants sat in a small room in front of a color monitor, speakers and a PsyScope button box. Experimenters read instructions to the participants before baseline, practice and experimental blocks. The baseline measure, which provided data on participants' response rate to simple visual stimuli, consisted of 30 presentations of a line-drawn face to either side of the monitor (following a warning beep). Participants indicated where the face appeared on the screen by using their right index finger to press the left or right button on the button box as quickly as possible.

After completing the baseline task, task familiarization and practice blocks of three and six trials were run, followed by two experimental blocks of 48 trials each, with a rest period between the two latter blocks. A trial consisted of the following: After a warning beep, drawings of two animals were projected on the left and right sides of the monitor over a gray background. The nouns referring to the animals were heard in succession (to unambiguously identify the drawings), followed by presentation of a sentence involving both animals. Participants were instructed to use their right index finger in order to press the button corresponding to the picture of the animal doing the bad action; the picture chosen by the participant was briefly highlighted before the screen was reset for the next trial. Each trial was cued up by the experimenter, who observed the subjects' performance and demeanor to assure that they were remaining attentive and alert. (If the participant did not respond by 5000 milliseconds after offset of the sentence, the next trial would start automatically.) Experimenters also provided encouragement to participants, and granted short breaks to those who appeared to be losing focus.

Order of visual and auditory stimuli presentation was fully randomized for each participant, as was presentation of trials. It was emphasized that participants should attempt to respond as accurately and quickly as possible to the stimuli. Accuracy feedback was not provided.

Results

We report results in the following order: (1) Typically developing children (TD) only; (2a) Children with language impairment (LI) alone; (2b) Children with LI compared to age-matched TD children; (2c) Children with LI compared with the youngest TD children; (3a) Children with focal lesions (FL) alone; (3b) Children with FL compared to age-matched TD children; (3c) Children with FL compared with the youngest TD children; (4) Children with FL compared to age-matched children with LI. For all sections, we report analyses of variance (ANOVAs) for Percent Correct Response (CR) and Corrected Mean RT on accurate responses only (RT), with Sentence Type (Active, Subject Cleft, Object Cleft, Passive) and Subject-Verb Agreement Cue (Present, Absent) as within-subjects factors. We do not report main effects or interactions for solely within-subjects factors in our cross-population analyses, as these are redundant with single-population analyses and not meaningful theoretically.

Table 3 *Number of participants per age (rounded to nearest year) for each group*

Age	FL	LI	TD
5	0	0	5
6	0	0	5
7	6	6	9
8	1	7	10
9	1	3	10
10	2	2	10
11	1	1	4
12	5	2	12
13	0	1	15
14	2	0	11
15	0	2	6
16	1	0	5
17	0	0	1
18	1	0	0
Total	20	24	103

In order to reduce the influence of outliers on age-related analyses, we assigned typically developing subjects to two- or three-year 'bins' based on their age rounded to the nearest year; these were 5–6, 7–8, 9–10, 11–12, 13–14 and 15–17. For our clinical populations, we created two broader age bins (5–9, 10–18) as the number of subjects per age was small and unevenly distributed. (The numbers of subjects per age and group are shown in Table 3.) For cross-population analyses, we paired each clinical subject with an age- and gender-matched counterpart (see below). We carried out all ANOVAs with SuperAnova and Statview 5.0 packages for Macintosh. *P*-values reported for all within-subjects factors are Geisser-Greenhouse (G-G) corrected (Geisser & Greenhouse, 1958), and all analyses used subjects as the random factor, as sentence items are extremely homogeneous (Clark, 1973). Reported means and differences of means were rounded to the nearest integer value in percent (accuracy) or milliseconds (reaction time). Cell means for all conditions can be viewed at http://crl.ucsd.edu/~fdick/devcsi/means.html.

(1) Typically developing children – main effects

In order to gain a global understanding of the effects and interactions of Age, Sentence Type and Agreement Cue, we performed omnibus ANOVAs with mean correct response (CR) and mean corrected reaction time (RT) as dependent variables. As would be predicted from countless developmental studies, CR and RT were modulated by age (CR: $F(5, 97) = 3.485$, $p = .0061$; RT: $F(5, 95) =$

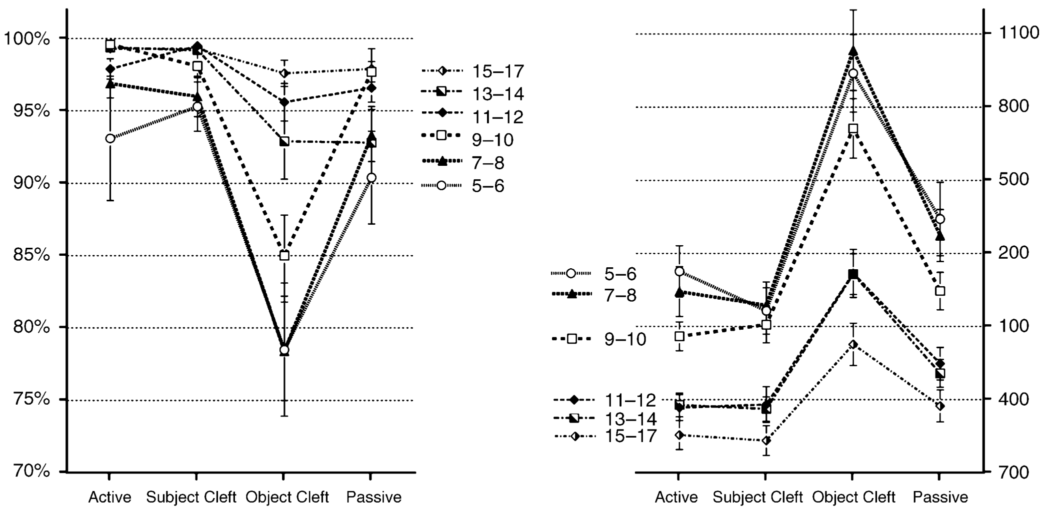


Figure 1 *Age group by sentence type effects with accuracy and corrected reaction time measures for typically developing children.*

7.806, $p = .0001$);[1] here, the older the child, the more accurate the response and the lower the reaction time. A Fisher's Protected Least Significant Difference test (groupwise controlled at the .05 level) showed that children ages 5–8 were significantly less accurate than children 9 and older (no other contrasts were significant); corresponding analyses of RTs showed a similar quasi-'stepwise' function, but one 'delayed' in developmental time, where correct reaction times of children 11–17 did not differ from each other, but were significantly lower than all children ages 10 and younger (who also did not significantly differ from each other).

Both Sentence Type and Agreement had overall effects on accuracy and reaction time. The effect of Sentence Type (CR: $F(3, 291) = 26.024$, $p = .0001$; RT: $F(3, 285) = 164.865$, $p = .0001$) was again as in previous studies of syntactic processing, where sentences with canonical word order (Actives, Subject Clefts) were interpreted more accurately and quickly than sentences with non-canonical word order (Passives, Object Clefts). Within-subjects G-G corrected pairwise comparisons ($p < .05$) showed a profile of accuracy such that CRs for Actives = Subject Clefts > Passives > Object Clefts; the same hierarchy of difficulty was seen in reaction times, where RTs for Actives = Subject Clefts < Passives < Object Clefts. The presence of disambiguating noun-verb agreement information also improved children's overall accuracy – (CR: $F(1, 97) = 7.760$, $p = .0064$) – albeit only by 1%. Agreement cues did not significantly affect overall reaction times.

[1] Differences between degrees of freedom for equivalent tests in CR and RT are due to the elimination of small numbers of subjects for RT calculations; this occurs because RTs are based on correct responses only, hence subjects with no accurate responses in a particular cell will be eliminated from analyses. For comparisons between clinical populations and age-matched typically developing children, both the LI or FL child with missing cells and their matched control were excluded from analyses.

Typically developing children – first- and second-order interactions

A significant Sentence Type × Agreement interaction over accuracy (CR: $F(3, 291) = 13.731$, $p = .0001$) showed that disambiguating noun-verb agreement information helped raise accuracy for the most difficult sentences, object clefts, from 86% (no cue) to 91% (with cue). Sub-ANOVAs performed on each sentence type, with Agreement as the within-subjects variable, showed that accuracy only significantly differed with agreement on Object Clefts (CR: $F(1, 102) = 18.542$, $p = .0001$; all other sentence types $p > .05$). There was no significant reciprocal Sentence Type × Agreement in reaction times, suggesting that collapsed over age, typically developing children could use the agreement cue to help interpret object clefts without requiring additional processing time.

Chronological age also differentially modulated children's interpretation of sentence types, as shown by a significant interaction of Sentence Type and Age Group for both accuracy and reaction time (CR: $F(15, 291) = 2.882$, $p = .0038$; RT: $F(15, 285) = 3.746$, $p = .0003$ – see Figure 1). Here, we see performance accuracy on actives and subject clefts converging towards ceiling levels early on, with 5–6-year-olds at 93% and 95% on actives and subject clefts, respectively; by 9–10 years of age, performance is at 99% and 98%. Passives lag behind somewhat, with the youngest children performing at 90% correct, with a slow climb towards asymptote in the oldest children. Object clefts show the longest and steepest developmental trajectory, with no improvement in accuracy from ages 5–8 (at a level of 78%), followed by incremental jumps in each age group to adult levels. Sub-ANOVAs (Sentence Type × Age Group Pair) comparing each age group to the next oldest (e.g. 5–6 to 7–8, 7–8 to 9–10) showed that the greatest shift in comprehension profiles occurred between the 9–10 and 11–12 age group (CR: $F(3, 102) = 5.658$, $p = .0167$); no other 4-year period of development showed a significant interaction with Sentence Type. By the ages of 15–17, children appear to have reached adult levels of accuracy; when we perform the same sub-ANOVA comparing children 15–17 to the group of 25 young adults reported in Dick *et al.* (2001) who underwent an almost identical experimental paradigm, we see no main effect of age or significant interaction with sentence type in accuracy scores.

Reaction time data generally mirror those for accuracy, although the developmental trajectory is somewhat more drawn out over time. Corrected reaction times for actives and subject clefts remain relatively constant from ages 5–10, while reaction times for passives (200–250 milliseconds slower than actives/subject clefts at ages 5–8) drop somewhat later in this age range. Reaction times for object clefts averaged almost a second longer than for subject clefts and actives between ages 5–8, and began to drop slightly between ages 9–10. As in the accuracy data, there was a significant change in reaction times between ages 9–12, with all reaction times dropping considerably (as seen in the pairwise comparisons for Age Group). Reaction times for object clefts descended most precipitously, as reflected by a marginally significant interaction in the sub-ANOVA of Age Group Pair × Sentence Type for the 9–12 age contrast (RT: $F(3, 102) = 2.857$, GG-corrected $p = .0745$). We see a continuation of this trend through to young adults; when we compare reaction times in age groups 13 and older – including the young college students from Dick *et al.* (2001) – we again find a significant Sentence Type × Age Group interaction (RT: $F(6, 177) = 5.205$, $p = .0009$), where reaction times on passives and especially object clefts become progressively faster relative to the quasi-asymptote observed for actives and subject clefts.

The impact of the presence or absence of agreement cues on accuracy (but not reaction times) also changed with age (CR: $F(5, 97) = 2.431$, $p = .0403$). Within-Age-Group ANOVAs showed a non-significant trend toward an increase in accuracy with an agreement cue in ages 5–6 (CR: $F(1, 27) = 3.126$, $p = .1109$), a more robust increase in accuracy with agreement between ages 9–10 (CR: $F(1, 57) = 8.515$, $p = .0088$), and no other significant effects of agreement within other ages (even when groups were combined with the next-closest age to increase number of subjects in a cell). It is possible that children in these two age groups may need the additional 'leg-up' provided by agreement in order to better interpret the sentence types that are most challenging for them.

The significant Age Group × Sentence Type × Agreement for accuracy data only (CR: $F(15, 291) = 2.086$, $p = .0178$) showed that object cleft performance was significantly improved when an agreement cue was available, particularly in the 5–6 and 9–10 year olds. Sub-ANOVAs within each age bin showed significant interactions of Sentence Type × Agreement for both these groups (CR: $F(3, 27) = 4.919$, $p = .0164$; $F(3, 57) = 4.906$, $p = .0192$, respectively), where object cleft performance was markedly increased with the help of agreement (15% in 5–6-year-olds, 8% in 9–10-year-olds). Children 11 and up showed this increased accuracy with object clefts, but to a much more limited degree; the Sentence Type × Agreement interaction only reached significance when two or more age groups were combined, and was most evident when data from all children 11 and older were merged (CR: $F(3, 153) = 5.677$, $p = .0027$). (The 7–8-year-old group had a small mean numerical increase in object cleft accuracy with agreement cue, but this was overwhelmed by within-group variance.) Interestingly, we also observed a *decrease* in accuracy on passives with an agreement cue for older children ages 11–17 ($p = .0190$), but not younger children. This effect may be due to the highlighting of the grammatical subject by the agreement cue; unlike in other types of sentences, the subject is not the sentential agent, thus providing a somewhat 'misleading' clue to agency. In other words, older children may associate the agreement cue with agency (a semantic category) rather than subjecthood (an abstract grammatical role). Interestingly, there is no sign of this effect in college students (Dick *et al.*, 2001), again suggesting that the process of refining cue strengths is one that continues up until early adulthood.

(2a) Children with language impairment

As with the typically developing children, we performed an omnibus 2-within (Sentence Type × Agreement), 1-between (Age Group) ANOVA on both accuracy and corrected reaction time measures. (Recall that age bins for clinical populations are coarser than those for TD children, i.e. 5–9 and 10–18 years old.) Here, language-impaired children's accuracy and reaction times varied significantly over sentence type (CR: $F(3, 66) = 18.710$, $p = .0001$; RT: $F(3, 66) = 21.317$, $p = .0001$); GG-corrected pairwise contrasts ($p < .05$) between sentence types showed a profile of accuracy such that Actives = Subject Clefts > Passives > Object Clefts; reaction times again mirrored this profile, where Actives = Subject Clefts < Passives < Object Clefts.

In contrast with our findings for typically developing controls, LI children's overall accuracy did not change significantly in the presence of agreement. Nor was there a significant change in accuracy or reaction time with age group. It is important to note that any lack of age-related effects may be due to the lack of statistical power to detect them, due both to the small sample size and unequal distribution of ages.

Turning to some of the interactions, the presence of an agreement cue differentially affected LI children's performance over sentence types, as reflected in a Sentence Type × Agreement interaction for both accuracy (CR: $F(3, 66) = 2.820$, $p = .0532$ – marginal) and reaction time (RT: $F(3, 66) = 8.652$, $p = .0003$). As with typically developing children, LI children were more accurate in comprehending object cleft sentences with an agreement cue (effect of agreement cue within object clefts only, ($F(1, 23) = 7.163$, $p = .0135$, not significant within any other sentence types)). However, this increased accuracy came at the cost of a reduced speed of processing, where object cleft sentences with agreement cues took 362 milliseconds longer to respond to than did those without agreement cues (agreement effect within object clefts only, ($F(1, 22) = 8.360$, $p = .0085$)). In other words, children with LI are able to use agreement information to improve performance on the most difficult sentence types, like typically developing children; however, unlike typically developing children, the children with LI pay a price in reaction times for this use of agreement information. In contrast, reaction times in the much easier subject cleft sentences were *decreased* in the presence of an agreement cue ($F(1, 23) = 4.299$, $p = .0495$), as were reaction times for passives at a marginal level of significance ($F(1, 21) = 4.164$, $p = .0541$). In line with a host of studies indicating that agreement morphology is challenging for children with LI, we suggest that use of the agreement cue is relatively costly for these children, compared with typically developing controls. For 'easier' sentences, the cue can be used in the normal way to speed up reaction times (at least for those sentences on which the children were accurate), but when the same agreement cue is used to interpret the 'hardest' sentence stimuli (i.e. the object clefts), children with LI have to slow down markedly to put these sources of information together.

No other effects or interactions within the language-impaired group approached statistical significance.

(2b) Language-impaired children compared to age- and gender-matched typically developing children; group effects and interactions

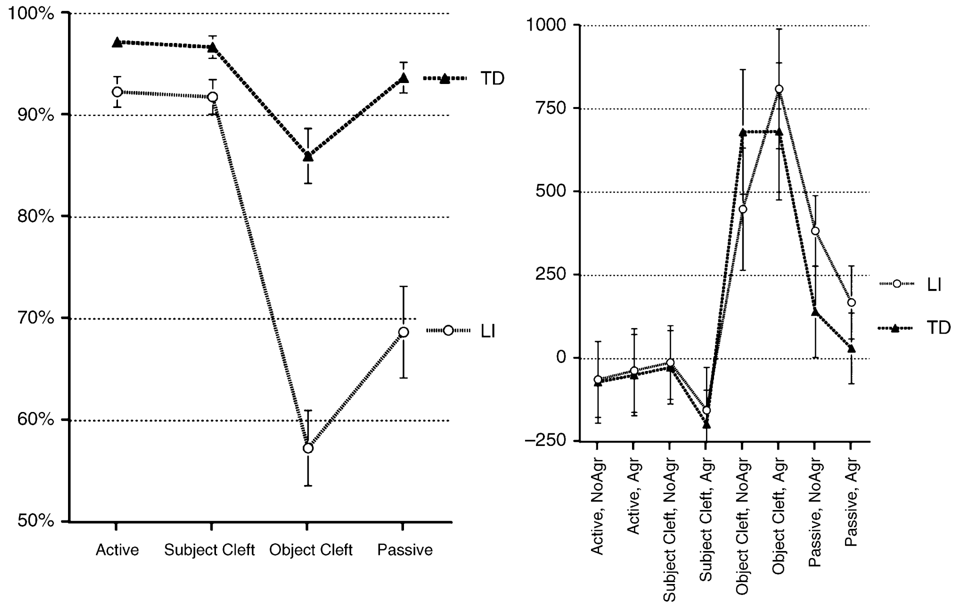
Before beginning statistical analyses, we matched each of our language-impaired children to a typically developing (TD) child from our sample. If there was more than one TD child who was an exact gender and age match (within 1 month), we randomly drew the match from that subsample; otherwise we selected the TD child who was closest in age and matching in gender when possible.

With this balancing for age and gender, the omnibus 2-within- (Sentence Type × Agreement), 1-between-subject (Group (LI vs. TD)) ANOVA showed a main effect of Subject Group, where LI children were less accurate overall (by 16%) than were their matched controls (CR: $F(1, 46) = 24.139$, $p = .0001$). However, there was no significant overall difference in reaction times between the two groups (keeping in mind that RTs were analyzed only for those sentences on which a correct response was given). LI children and their matched controls also differed in their accuracy over sentence type (CR: $F(3, 138) = 9.893$, $p = .0002$), where LI children were vastly less accurate in comprehending passives and especially object clefts than were matched TD children (LI %CR – TD %CR = 25% passives and 29% object clefts), while this gulf in accuracy narrowed to just a few percent for actives and subject clefts (5% group difference for both actives and subject clefts – see Figure 2a). Again, this disparity in accuracy across sentence types was not reflected in reaction time measures (i.e. no significant group by sentence type interaction).

Reaction time data did reveal one significant effect in this group comparison, namely a significant three-way interaction between Sentence Type, Agreement and Group (RT: $F(3, 126) = 3.062$, $p = .0362$). This effect was driven solely by the different patterns of LI and TD children's reaction times for object clefts with and without agreement cues (see Figure 2b). Here, LI children were 362 milliseconds *slower* with the agreement cue than without it, while TD children responded equally quickly (within 33 milliseconds) regardless of the presence or absence of an agreement cue. This interaction reflects the situation described earlier in analyses of the LI group only: children with LI are able to use agreement morphology to improve their interpretation of 'hard' sentences, but use of this cue is more costly for them (in terms of processing speed) than it is for typically developing controls matched for age and gender.

(2c) Comparison of LI children to youngest TD children

As noted in the introduction, LI children are often characterized as showing a 'delayed' profile of language development. Thus, it might be informative to directly compare the performance of our sample of LI children to that of the youngest TD children to see if their comprehension profiles are similar or divergent. A 2-within (Sentence Type, Agreement) × 1-between (Group) ANOVA comparing all LI children to TD children ages 5–7 showed not only that LI children were less accurate than



Figures 2a and 2b *Sentence type by child group effects with accuracy (2a) and sentence type by agreement by child group with corrected reaction time (2b) measures for children with language impairment and age-matched typically developing children.*

the youngest children (CR: $F(1, 41) = 10.033$, $p = .0029$), but that LI children were considerably less accurate than young TD children on object clefts and passives, reflected in the Sentence Type × Group interaction (CR: $F(3, 123) = 5.963$, $p = .0041$) as well as in cross-group comparisons over sentence type (Object Cleft CR: $F(1, 41) = 9.818$, $p = .0032$; Passive CR: $F(1, 41) = 9.124$, $p = .0043$; actives and subject clefts did not differ). These significant accuracy differences in passives and object clefts held even when we further restricted TD age range to 5–6 years or even just 5 years of age (although in the latter case the TD > LI passive difference was marginal). This differential accuracy over sentence type also holds to some degree when we compare only the oldest (10–17) LI to the youngest (5–7) TD children (CR: $F(1, 25) = 2.909$, $p = .0704$), with a significant TD > LI advantage on object cleft accuracy – the TD > LI difference for passives is numerical only. (No other accuracy differences between these subgroups were significant, nor were there any significant effects on RT.) In short, these LI children appear to have serious and long-lasting problems in comprehending non-canonical word orders, even more so than the youngest children in our sample. This is the case even for the older LI children, although we hasten to add that only five of these children are ages 12 and above.

(3a) Children with focal lesions

With this group of children, we had an additional question to ask: does laterality of lesion differentially affect accuracy and reaction time? Although there was no significant difference between the LH and RH groups in age (LH mean age 10.4 (SD 3.3), RH mean 10.9 (SD 3.5)), children with right- and left-hemisphere damage were not evenly distributed over age bins. For this reason, we ran two separate ANOVAs, with either Hemisphere (LH vs. RH) or Age Group (5–9 vs. 10–18) as the between-subjects variable, and (as before) Sentence Type and Agreement Cue as the within-subjects variables.

The Sentence Type × Agreement × Hemisphere ANOVA revealed a main effect of Sentence Type, similar to results observed with our other groups of children. Sentence Type modulated FL children's accuracy (CR: $F(3, 54) = 21.599$, $p = .0001$) and corrected reaction time (RT: $F(3, 54) = 29.478$, $p = .0001$); GG-corrected pairwise comparisons ($p < .05$ unless otherwise noted) showed a profile of accuracy such that Actives = Subject Clefts > Passives > Object Clefts, again mirrored in reaction times, such that Subject Clefts < Actives < Passives < Object Clefts. (Accuracy differences between Actives/Subject Clefts and Passives were marginally significant, with $p = .0669$ and $p = .0774$, respectively; reaction times for subject clefts were marginally faster than for actives, $p = .0918$.)

FL children's accuracy on each sentence type was modulated by agreement similarly to our other child groups, albeit at marginally significant levels (CR: $F(3, 54) = 3.034$, $p = .0586$); as before, accuracy on object clefts was higher with an agreement cue, whereas passives *without* an agreement cue were comprehended more accurately. As noted above, we interpret the negative effects of agreement on passives to indicate that children associate the agreement cue with agency rather than the subject role.

With one exception, there were no effects of Hemisphere on performance; the only significant effect involving laterality of lesion was an interaction of Hemisphere and Agreement in reaction times (RT: $F(1, 18) = 7.735$, $p = .0123$), where right- and left-lesioned children's reaction times to sentences *with* agreement cues were virtually identical, but reaction times without agreement diverged considerably according to lesion side. The direction of this result was unanticipated: children with lesions to the left hemisphere (LHD) responded 267 milliseconds *faster* in the absence of agreement cues, compared to children with lesions in the right hemisphere (RHD). The absence of lesion-side effects is no longer surprising to us, because many years of research in our laboratories and those of other investigators have repeatedly shown equivalent performance on language tasks by 5–7 years of age in children with left- vs. right-hemisphere injury (Bates & Roe, 2001; Vargha-Khadem, Isaacs & Muter, 1994). However, the slower performance by children with RHD in the absence of agreement cues that we have uncovered here is surprising, in light not only of our own prior findings but also in comparison with a host of studies of language performance in adults with RHD vs. LHD. At this point we can only speculate about the basis for this result, noting that children with RHD sometimes show limitations in aspects of attention in both verbal and non-verbal tasks (Moses & Stiles, 2002). It may be the case that the sentences without agreement require more attention, resulting in slower performance for children who have a harder time concentrating on the task.

When we performed the second 2-within-, 1-between-subjects ANOVA, with Age Group as the between-groups factor, we found no significant simple or compound effects of age in these children on accuracy. However, we did find a significant interaction of Age Group × Sentence Type on reaction times (RT: $F(3, 54) = 4.338$, $p = 0203$), where older FL children are numerically faster in responding to object clefts and passives (neither of the group contrasts within sentence type are significant, however, with p-values at .10 and .21, respectively). This effect suggests that processing of these sentence types is becoming slightly easier for FL children with increasing experience.

(3b) Children with focal lesions compared to age- and gender-matched typically developing children; group effects and interactions

Before comparing these two populations, we used the method described in (2b) to form a group of age- and gender-

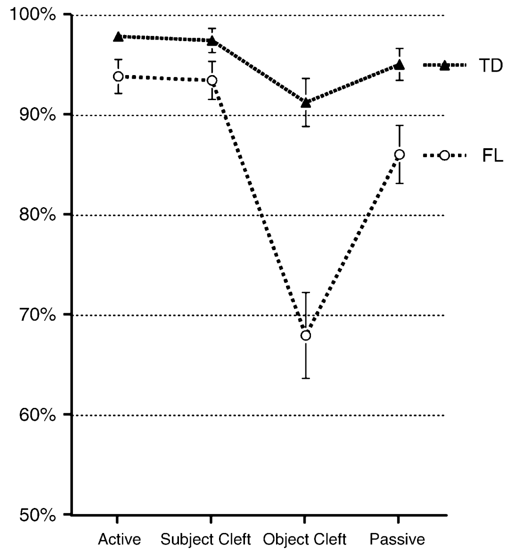


Figure 3 *Sentence type by child group effects on accuracy for children with focal lesions and age-matched typically developing children.*

matched typically developing children. As with our other group comparisons, we then performed a 2-within-(Sentence Type × Agreement), 1-between-subjects (Group) ANOVA on both accuracy and reaction time. Here, children with FL were not only less accurate overall (as shown by a main effect of Group on accuracy (CR: $F(1, 38) = 7.596$, $p = .0089$)) but were differentially sensitive to sentence type relative to their matched peers, as revealed in a Sentence Type × Group interaction (CR: $F(3, 114) = 10.175$, $p = .0006$ – see Figure 3). Specifically, children with FL were almost as accurate as their matched controls in interpreting actives and subject clefts (4% less accurate for both sentence types), but were considerably less accurate on passives (9% difference) and especially object clefts (23% difference) – see Figure 3. This effect was mirrored in the reaction time measures with a marginal Sentence Type × Group interaction (RT: $F(3, 114) = 2.348$, $p = .1016$) where small differences between group mean reaction times for actives and subject clefts (190 and 82 milliseconds, respectively) contrasted with larger differences for passives and object clefts (357 and 310 msec). No other effects involving group were close to statistical threshold.

(3c) Comparison of FL children to youngest TD children

Here, we again compare our clinical population sample to the youngest typically developing children (ages 5–7). Unlike the parallel contrast between LI and young TD children, there were no group-based main effects or interactions, providing some evidence that this group of FL children tend to perform similarly to younger healthy children, unlike their LI counterparts. It is worth noting that there is a significant Group × Sentence Type on accuracy scores (where younger FL children perform worse on object clefts and passives) when the oldest four FL children (ages 14, 14, 16 and 18) are excluded from analyses, suggesting that the developmental trajectory for FL children is extremely protracted.

(4) FL children compared to age-matched LI children

Because our groups of FL and LI children were small and unevenly distributed over age, we were unable to match all our children with focal lesions to language-impaired children. In order to preclude any between-group age confound, we restricted our analyses to these 15 matched FL children and their age-matched LI complements. Here, we found no reliable effects or interactions with Group in the 2-within-, 1-between-subjects ANOVA, despite relatively large numerical differences between the two groups. That is, the children with FL performed numerically better than age- and gender-matched children with LI, but this trend was buried by large individual differences within each group.

We were concerned that possible group differences might be masked by the diminished sample size, so we also ran an additional group analysis including all FL and LI children, controlling for age differences by entering age as a continuous covariate. This 2-within (Sentence Type, Agreement), 2-between (Group, Age) ANOVA again showed no significant differences between LI and FL groups, but did show that age was positively correlated with overall accuracy when controlling for overall group differences (CR: $F(1, 41) = 4.700$, $p = .0360$). (The age effect holds even when the high-leverage 16- and 18-year-old FL subjects are excluded – $p = .0547$; a non-parametric test (Kendall's Tau) also showed a significant ($p = .0363$) age-accuracy correlation.) This result suggests that there is some overall improvement in sentence comprehension accuracy in our clinical groups, but that it can only be detected when sample size is sufficiently large.

Summary and discussion

Among the more important results of the present study is a confirmation of the finding that language comprehension skills in typically developing children increase in speed and accuracy across the first two decades of life, especially for sentence types that are low in frequency and word order regularity (e.g. passives and especially object clefts). This kind of gradual, probabilistic development has been observed by others (going back to Slobin's (1966) early studies of passives and negatives), but is confirmed here in a tightly controlled on-line paradigm with millisecond resolution. These results are compatible with the predictions of the Competition Model, or other connectionist-style constraint-satisfaction models in which 'knowledge' involves the gradual strengthening of representations over time, driven by statistical properties of the input, interacting with cue validity and cue cost.

The nature of the sentence type effects observed here is also compatible with models in which whole sentence types like clefts and passives are represented in a distributed fashion, so that variations in frequency and validity apply in the 'microstructure' of these sentence types. Subject clefts are actually extremely low in frequency (Roland, Dick & Elman, submitted), but they bear a strong similarity to the highly frequent active sentence frame, in that the two structures share the same word order (Agent-Verb-Object) and similar morphology. Object clefts are not only low in frequency, but they also carry low-frequency word order within their microstructure (Object-Agent-Verb). Passives also have low frequency word order (Object-Verb-Agent), and atypical morphology (verbs in the participial form; presence of a by-phrase), and yet they fare better than object clefts in our experiment. We suggest this is the case because passive morphology is both perceptually salient and high in cue validity, compensating partially for their low-frequency word order. (Note, however, that passives are considerably more frequent than object clefts, both in raw token frequency as well as the frequency of the underlying word orders – Roland, Dick & Elman, submitted.) The development of sentence comprehension skills appears to mirror these distributional facts, suggesting that representations are strengthened over development not just at the level of the whole sentence type, but at the level of sentential microstructure, where principles of frequency and regularity, cue validity and cue cost apply.

The effects of subject-verb agreement in the present study were small, as we might expect in a study of English, a language in which agreement morphology is relatively low in cue validity, and is subordinated by word order throughout the listener's language-learning history. Although these English listeners were able to use the agreement contrast to facilitate response, this effect was due almost entirely to performance on the more difficult object cleft sentence type. Furthermore, the 'leg up' provided by morphology was actually more evident in the younger children, compensating for the relative weakness of object clefts in this group. This is an example of a 'coalition of weak cues', a phenomenon that is often observed in transitional periods of learning, before children reach an adult configuration of cue strengths (MacWhinney & Bates, 1989; Bates *et al.*, 2001). Finally, we note a peculiar but intriguing trend in the effect of subject-verb agreement on passives, where presence of an agreement contrast actually seems to make things worse, resulting in a slight but significant decrease in accuracy.[2] We speculate that this occurs because children tie the agreement cue not to the abstract subject role (which is carried by the patient of the action in a passive sentence) but to the agent role that typically holds the subject role in high-frequency active sentence types.

Turning to results for our two clinical groups, in almost all our studies to date we have found the children with a history of early focal brain injury outperform children with LI on language tasks (see papers in Wulfeck & Reilly, in press). Results of the current study represent an interesting exception to this trend. Although the FL group performed numerically higher than matched children with LI, there were no significant effects of group in any of the direct FL/LI comparisons. (We should emphasize that this may stem at least in part from the low power (.14–.39) of group-based analyses due to the small sample sizes and within-group variability.) Comparisons between each clinical group and age- and gender-matched TD children showed that both FL and LI children find passive and especially object cleft sentences extremely difficult to process – much like adult aphasic patients and college students under certain stress conditions (Dick *et al.*, 2001). Comparisons between FL and the youngest TD children suggest that FL children's performance is not dissimilar to that of very young TD children; however, LI children's accuracy on passives and object clefts is significantly worse than even 5-year-old TD children. Because our LI sample includes only five children ages 12 and above, it is impossible to know whether this deficit in passive and object cleft comprehension represents a long and severe delay in development (as has generally been found – Leonard, 1998) or a truly deviant profile of comprehension. (The same sampling problem makes any lack of age effects equally difficult to interpret.)

One piece of evidence pointing to LI children showing protracted language delay, rather than deviance, is their pattern of performance when using noun-verb agreement cues. Children with LI clearly have the knowledge and capabilities required to use subject-verb agreement information, which facilitated their response on some sentence types (e.g. decreasing reaction times for subject clefts, and increasing accuracy on object clefts). However, it appears that the costs required to use agreement information are higher in children with LI, who slowed down significantly on object clefts with an agreement contrast, while achieving (and perhaps in order to achieve) a correct response. These results add an interesting spin to the literature attesting to the special vulnerability of grammatical morphology in children with LI (Marchman, Saccuman & Wulfeck, 2003). Contrary to claims that LI children's morphological deficits are evidence for missing grammatical knowledge or competence (Gopnik & Crago, 1991; van der Lely & Ullman, 2001), the LI children in our study are using morphological information to speed processing of 'easy' sentences (subject clefts) or to raise comprehension accuracy for 'hard' sentences (object clefts), albeit with a cost to speed of processing. This kind of result suggests resource trade-offs in an overloaded system with limited resources, in line with several different processing models of grammatical deficits in LI (Joanisse, Manis, Keating & Seidenberg, 2000; Leonard, Deevy, Miller, Rauf, Charest & Robert, 2003).

These particular processing cost effects appear to be specific to the LI group, not appearing either in typically developing children or in the FL sample, who behave roughly like delayed typical controls, as noted above. However, we did find one peculiar hemispheric result within the FL group. An ANOVA comparing children with left- vs. right-hemisphere damage yielded only one significant result, and that was in the opposite direction from standard predictions that deficits will be greater with LHD. Specifically, we found a significant interaction between Hemisphere and Agreement on reaction times, in which LHD and RHD performed equally on sentences *with* an agreement cue, but RHD children were significantly slower than LHD on sentences *without* an agreement cue. Although we can only speculate regarding the origins of this unpredicted effect, we suggest that it may have something to do with the deficits in attention that are often reported for patients with RHD – given converging information, children with LHD and RHD perform the same, but with less information the child with RHD may require more time to arrive at a solution. Although this result is interesting (and worthy of replication), perhaps the most important conclusion from the FL group is that, once again, we fail to find evidence for a specific disadvantage for language among children with early damage to the left hemisphere. This stands in stark contrast with the adult aphasia literature, in which difficulties with passives, object clefts and other non-canonical word orders is often viewed as a hallmark symptom of agrammatic aphasia, and processing of complex syntax is typically attributed to areas within the

[2] The single exception to this general observation regarding passives was the decrease in reaction times in the presence of an agreement cue in LI children.

left hemisphere (for a detailed discussion, see Dick *et al.*, 2001).

Taken together, results testify to the value of on-line methods in the study of language development beyond the preschool years in assessing changes in the temporal dynamics of sentence comprehension from kindergarten through adolescence. Results for the two clinical populations in our study suggest that these methods can also play an important role in characterizing the nature of deficits and delays in language use.

Acknowledgements

The research reported here was supported in part by funding from NIH-NINDS Grant P50 NS22343 to EB, as well as NIH Center for Research in Language Grant T32 DC00041 'Language, Communication, and the Brain' pre-doctoral fellowship to FD and NIH-NIMHG T32 MH20002-02 Institute for Neural Computation predoctoral fellowship to FD. We would like to thank the children and parents who participated in this study. Thanks also to our colleague, Doris Trauner, MD, who directs the medical/diagnostic core and to the research staff at the San Diego Project in Cognitive and Neural Development/Language Research Center for assistance in screening and testing. We also thank Bob Buffington and staff at UCSD's Center for Research in Language for technical assistance and computer programming. Finally, we are very grateful to our action editor and two anonymous reviewers for their detailed and helpful suggestions.

References

Abbate, M.S., & LaChapelle, N.B. (1984a). *Pictures, please! A language supplement*. Tucson, AZ: Communication Skill Builders, Inc.

Abbate, M.S., & LaChapelle, N.B. (1984b). *Pictures, please! An articulation supplement*. Tucson, AZ: Communication Skill Builders, Inc.

American Psychological Association (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, **47**, 1597–1611.

Bates, E., Devescovi, A., & Wulfeck, B. (2001). Psycholinguistics: a cross-language perspective. *Annual Review of Psychology*, **52**, 369–398.

Bates, E., & MacWhinney, B. (1987). Competition, variation and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Hillsdale, NJ: Erlbaum.

Bates, E., MacWhinney, B., Caselli, C., Devescovi, A., Natale, F., & Venza, V. (1984). A cross-linguistic study of the development of sentence interpretation strategies. *Child Development*, **55**, 341–354.

Bates, E., & Roe, K. (2001). Language development in children with unilateral brain injury. In C.A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 281–307). Cambridge, MA: MIT Press.

Bates, E., Wulfeck, B., & MacWhinney, B. (1991). Cross-linguistic research in aphasia: an overview. *Brain and Language*, **41** (2), 123–148.

Bever, T.G. (1970). The cognitive basis for linguistic structures. In J.R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York: John Wiley & Sons.

Blackwell, A., & Bates, E. (1995). Inducing agrammatic profiles in normals: evidence for the selective vulnerability of morphology under cognitive resource limitation. *Journal of Cognitive Neuroscience*, **7** (2), 228–257.

Bloom, L., & Lahey, M. (1978). *Language development and language disorders*. New York: John Wiley & Sons.

Bloom, L., Lahey, M., Hood, L., Lifter, K., & Fiess, K. (1980). Complex sentences: acquisition of syntactic connectives and the semantic relations they encode. *Journal of Child Language*, **7** (2), 235–261.

Bloom, L., Miller, P., & Hood, L. (1975). Variation and reduction as aspects of competence in language development. In A. Pick (Ed.), *Minnesota symposium on child psychology* (Vol. 9, pp. 3–55). Minneapolis: University of Minnesota Press.

Booth, J., MacWhinney, B., & Harasaki, Y. (2000). Developmental differences in visual and auditory processing of complex sentences. *Child Development*, **71** (4), 981–1003.

Booth, J., MacWhinney, B., Thulborn, K., Sacco, K., Voyvodic, J., & Feldman, H. (2000). Developmental and lesion effects in brain activation during sentence comprehension and mental rotation. *Developmental Neuropsychology*, **18** (2), 139–169.

Brown, R. (1973). *A first language*. Cambridge: Harvard University Press.

Cedergren, H.J., & Sankoff, D. (1974). Performance as a statistical reflection of competence. *Language*, **50**, 333–355.

Clark, 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**, 335–359.

Cohen, J.D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: a new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers*, **25** (2), 257–271.

Demuth, K. (1989). Maturation and the acquisition of Sesotho passive. *Language*, **65** (5), 56–80.

Devescovi, A., D'Amico, S., & Gentile, P. (1999). The development of sentence comprehension in Italian: a reaction time study. *First Language*, **19** (56, pt 2), 129–163.

Dick, F., Bates, E., Wulfeck, B., Utman, J.A., Dronkers, N., & Gernsbacher, M.A. (2001). Language deficits, localization, and grammar: evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychological Review*, **108** (4), 759–788.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, **14**, 179–211.

Feldman, H., MacWhinney, B., & Sacco, K. (2002). Sentence processing in children with unilateral brain injury. *Brain and Language*, **83** (2), 335–352.

Fletcher, P., & MacWhinney, B. (Eds.) (1995). *The handbook of child language*. New York: Cambridge University Press.
Fodor, J.A., Bever, T.G., & Garrett, M.F. (1974). *The psychology of language*. New York: McGraw-Hill.
Geisser, S., & Greenhouse, S.W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. *Annals of Mathematical Statistics*, **29**, 885–891.
Gopnik, M., & Crago, M.B. (1991). Familial aggregation of a developmental language disorder. *Cognition*, **39** (1), 1–50.
Hayes, J.R. (1970). *Cognition and the development of language*. New York: John Wiley & Sons.
Joanisse, M., Manis, F., Keating, P., & Seidenberg, M. (2000). Language deficits in dyslexic children: speech perception, phonology, and morphology. *Journal of Experimental Child Psychology*, **77** (1), 30–60.
Kempe, V., & MacWhinney, B. (1999). Processing of morphological and semantic cues in Russian and German. *Language and Cognitive Processes*, **14** (2), 129–171.
Kilborn, K. (1991). Selective impairment of grammatical morphology due to induced stress in normal listeners: implications for aphasia. *Brain and Language*, **41** (2), 275–288.
Labov, W. (1969). Contraction, deletion and inherent variability of the English copula. *Language*, **45**, 715–762.
Leonard, L., Deevy, P., Miller, C., Rauf, L., Charest, M., & Robert, K. (2003). Surface forms and grammatical functions: past tense and passive participle use by children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, **46** (1), 43–55.
Leonard, L.B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.
MacDonald, M.C. (1999). Distributional information in language comprehension, production, and acquisition: three puzzles and a moral. In E. Brian MacWhinney (Ed.), *The emergence of language* (pp. 177–196). Mahwah, NJ: Lawrence Erlbaum.
MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Erlbaum.
MacWhinney, B., & Bates, E. (1989). *Cross-linguistic study of sentence processing*. New York: Cambridge University Press.
MacWhinney, B., Osman-Sagi, J., & Slobin, D. (1991). Sentence comprehension in aphasia in two clear case-marking languages. *Brain and Language*, **41** (2), 234–249.
MacWhinney, B., & Pleh, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition*, **29** (2), 95–141.
Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.
Marchman, V., Saccuman, C., & Wulfeck, B. (2003). Productive use of the English past tense in children with focal brain injury and specific language impairment. *Brain and Language*, Advance online publication.
Moses, P., & Stiles, J. (2002). The lesion methodology: contrasting views from adult and child studies. *Developmental Psychobiology*, **40** (3), 266–277.
Owens, R.E. (1991). *Language development: An introduction* (3rd edn.). New York: Merrill.
Reich, P.A. (1986). *Language development*. Englewood Cliffs, NJ: Prentice Hall.
Reilly, J., Bates, E., & Marchman, V. (2003). Narrative discourse in children with early focal brain injury. *Brain and Language*, advance online publication.
Reilly, J., Losh, M., Bellugi, U., & Wulfeck, B. (2003). 'Frog, where are you?' Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and Language*, advance online publication.
Roe, K. (2003). Working memory and sentence comprehension in adults and children. Unpublished dissertation.
Roland, D., Dick, F., & Elman, J. (submitted). Frequency of basic English grammatical structures: a corpus analysis.
Rumelhart, D.E., & McClelland, J.L. (1986). *Parallel distributed processing: Volume 1: Foundations*. Cambridge, MA: MIT Press.
Slobin, D.I. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, **5**, 219–227.
Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, **10**, 174–215.
Tager-Flusberg, H., & Cooper, J. (1999). Present and future possibilities for defining a phenotype for specific language impairment. *Journal of Speech, Language and Hearing Research*, **42** (5), 1275–1278.
Turner, E.A., & Rommetveit, R. (1968). Focus of attention in recall of active and passive sentences. *Journal of Verbal Learning and Verbal Behavior*, **7**, 543–548.
Tyler, L.K., & Marslen-Wilson, W.D. (1981). Children's processing of spoken language. *Journal of Verbal Learning and Verbal Behavior*, **20**, 400–416.
van der Lely, H.K., Rosen, S., & McClelland, A. (1998). Evidence for a grammar-specific deficit in children. *Current Biology*, **8** (23), 1253–1258.
van der Lely, H.K.J., & Ullman, M.T. (2001). Past tense morphology in specifically language impaired and normally developing children. *Journal of Language and Cognitive Processes*, **16** (2), 177–217.
Vargha-Khadem, F., Isaacs, E., & Muter, V. (1994). A review of cognitive outcome after unilateral lesions sustained during childhood. *Journal of Child Neurology*, **9** (2), 67–73.
Von Berger, E., Wulfeck, B., Bates, E., & Fink, N. (1996). Developmental changes in real-time sentence processing. *First Language*, **16**, 193–222.
Weckerly, J., Wulfeck, B., & Reilly, J. (2003). Morphosyntactic abilities in atypical populations: the acquisition of Tag Questions in children with early focal brain injury and children with specific language impairment. In B. Wulfeck & J. Reilly (Eds.), Plasticity and development: language in atypical children. Special issue, *Brain and Language*, Advance online publication.
Wiig, E.H., & Semel, E. (1984). *Language assessment and intervention for the learning disabled* (2nd edn.). Columbus, OH: Merrill.
Wulfeck, B.B. (1993). A reaction time study of grammaticality judgments in children. *Journal of Speech and Hearing Research*, **36**, 1208–1215.

Wulfeck, B., Bates, E., Krupa-Kwiatkowski, M., & Saltzman, D. (2003). On-line grammaticality sensitivity in children with early focal brain injury and specific language impairment. In B. Wulfeck & J. Reilly (Eds.), Plasticity and development: language in atypical children. Special issue, *Brain and Language*, Advance online publication.

Wulfeck, B., & Reilly, J. (Eds.) (in press). Plasticity and development: language in atypical children. Special issue, *Brain and Language*.

Received: 14 July 2003
Accepted: 2 November 2003