Morphosyntactic annotation of CHILDES transcripts*

KENJI SAGAE

Institute for Creative Technologies, University of Southern California

ERIC DAVIS

Language Technologies Institute, Carnegie Mellon University

ALON LAVIE

Language Technologies Institute, Carnegie Mellon University

BRIAN MACWHINNEY

Department of Psychology, Carnegie Mellon University

AND

SHULY WINTNER

Department of Computer Science, University of Haifa, Israel

ABSTRACT

Corpora of child language are essential for research in child language acquisition and psycholinguistics. Linguistic annotation of the corpora provides researchers with better means for exploring the development of grammatical constructions and their usage. We describe a project whose goal is to annotate the English section of the CHILDES database with grammatical relations in the form of labeled dependency structures. We have produced a corpus of over 18,800 utterances (approximately 65,000 words) with manually curated gold-standard grammatical relation annotations. Using this corpus, we have developed a highly accurate data-driven parser for the English CHILDES data, which we used to automatically annotate the remainder of the English section of CHILDES. We have also extended the parser to Spanish, and are

currently working on supporting more languages. The parser and the manually and automatically annotated data are freely available for research purposes.

INTRODUCTION

Explaining child language acquisition is one of the main challenges facing cognitive science, linguistics and psycholinguistics. Although all normal children succeed in learning their native tongue, neither psychology nor linguistics has yet succeeded in accounting for the many complexities of language learning. Within this general area, there has been particular attention to the acquisition of grammar, as it is expressed through morphosyntax, stimulated in large measure by Chomsky's theory of Universal Grammar and its attendant claims regarding innate principles and parameters. Beyond its theoretical importance, the measurement of morphosyntactic competence is crucial to applied work in fields such as developmental language disorders, schooling and literacy, and second language acquisition. To test and validate theoretical predictions quantitatively, researchers have increasingly come to rely on large corpora of transcript data of verbal interactions between children and parents to examine the development of morphosyntax. A standard source of data in this area is the CHILDES database (MacWhinney, 2000; http://childes.psy.cmu.edu), which provides 300 megabytes of transcript data for over twenty-five human languages, as well as a large amount of digitized audio and video linked to the transcripts.

There are now numerous studies that have used the CHILDES database to investigate the development of morphosyntax. However, these studies have typically been limited to using the database in its raw lexical form, without tags for part-of-speech, syntactic parses or predicate–argument information. Lacking this information, researchers have devoted long hours of manual analysis to locating and coding the sentences relevant to their hypotheses. Because these isolated manual annotation efforts of small segments of the database were not based on shared community standards, the resulting analyses are often discrepant and non-replicable, and much of the resulting corpus-based research on syntactic development has been only weakly cumulative. If syntactic parses were available, these analyses could be largely automated, allowing investigators to conduct a wider variety of tests in a more reliable fashion. Automatic syntactic analysis systems would also be of great value in clinical settings, allowing clinicians and clinical researchers to construct profiles of language delays by comparing small speech samples collected in structured interviews with a larger database of normed data. To this end, the raw information in the CHILDES corpora has been gradually enriched by a layer of morphological information. In particular, the English section of the database is augmented with

categorical and morphological information at the lexical level in the form of part-of-speech (POS) and morphological tags for each word. However, such information is usually insufficient for investigations dealing with the syntactic, semantic or pragmatic aspects of the data.

In this paper we describe an added layer of information, whereby the CHILDES database is annotated with utterance-level syntactic information, based on grammatical relations represented as labeled dependency structures. Although Sagae, Lavie & MacWhinney (2004) proposed an annotation scheme for syntactic information in CHILDES, until recently no significant amount of annotated data had been made publicly available. To correct this situation, we have developed a method for automatic annotation of the entire English CHILDES corpus (Eng-USA, Eng-UK, Clinical and Narrative), including utterances spoken by children and by adults. The first steps in this process involved manually annotating several thousands of words and continually revising and extending the annotation scheme during this process to account for the wide variety of structures found in real corpora. Next, we trained a state-of-the-art data-driven syntactic parser on our manually annotated corpus. We then further annotated a corpus of over 18,800 utterances (approximately 65,000 words), which we cross-checked manually to develop a gold standard. Of these 18,800 utterances, approximately 8,600 were spoken by children. We then retrained our statistical parser using this gold-standard corpus. In the final step, we used the parser to automatically annotate both the child and adult utterances in the entire English section of CHILDES. The gold-standard annotated data, the automatically annotated corpus and the parser can be freely downloaded from the CHILDES website.

This work has four major applications. First and foremost, we have now provided accurate information on 37 grammatical relations throughout the English section of CHILDES. Using these data, researchers can now directly track issues such as the development of the double-object construction, profiles of subject omission, the growth of verbal complements, quantifier scope, etc. In other words, these codes provide the basis for a wide range of automatic analyses of the growth of specific constructions or phrasal units. Second, these annotations now provide appropriate targets for the evaluation of competing models of syntactic development. To this end, we represent syntactic information using a scheme that focuses on specific grammatical relations, attempting to abstract away from theory-specific assumptions that may be adopted in different models of syntactic development. Third, the syntactic parser that produces these grammatical relation annotations, along with existing tools (MacWhinney, 2000; Parisse & Le Normand, 2000) for POS annotation, can be used to automate syntactic profiling through instruments such as Developmental Sentence Scoring (Lee, 1974) or the Index of Productive Syntax (Scarborough, 1990; Sagae, Lavie &

MacWhinney, 2005). Finally, this work with English can form the springboard for parallel work in other languages. Later, we will discuss how we have built a parallel system for Spanish, and some of the ways in which additional languages can provide specific new challenges for parser development.

APPLICATIONS

The syntactic annotation of child–adult linguistic interactions provides a valuable source of accurate information for researchers interested in language acquisition and development (MacWhinney, 2008), and the syntactic parser allows researchers to generate accurate annotations for new data. We outline some existing uses of the annotated corpus and syntactic parser.

Syntactic analysis of child language transcripts using a grammatical relation scheme has already been shown to be effective in a practical setting, namely in automatic measurement of syntactic development in children (Sagae *et al.*, 2005). That work relied on a phrase-structure statistical parser (Charniak, 2000) trained on newspaper articles, and the output of that parser had to be converted into grammatical relations, such as subject, object and adjunct. Despite the obvious disadvantage of using a parser trained on a completely different language genre, Sagae *et al.* demonstrated how current natural language processing techniques can be used effectively in child language work, achieving results that are close to those obtained by manual computation of syntactic development scores for child transcripts. Still, the use of tools not tailored for child language and extra effort necessary to make them work with community standards for child language transcription present a disincentive for child language researchers to incorporate automatic syntactic analysis into their work. We hope that the grammatical relations (GR) representation scheme and the parser presented here will make it possible and convenient for the child language community to take advantage of some of the recent developments in natural language parsing, as was the case with part-of-speech tagging when CHILDES specific tools were first made available.

A different area of research in which the syntactic annotation of CHILDES had been used is investigation of language acquisition processes, with an eye to determining the level of abstractness exhibited by early language. Borensztajn, Zuidema & Bod (2009) used the syntactically annotated Brown corpus in the framework of the data-oriented parsing (DOP) paradigm to measure the degree of abstractness in the speech of three children. In subsequent work, Bod (2009) used the UNSUPERVISED variant of DOP to simulate the linguistic capabilities of children (and adults). Here, the annotated corpus is not used for training (as the paradigm is unsupervised), but it IS used for evaluating the results of the learning algorithm.

708

Because reliable automatic syntactic annotation of CHILDES data at the utterance level depends heavily on accurate part-of-speech annotation at the word level, we introduced slight revisions to the existing part-of-speech annotation available in the English portion of CHILDES. These revisions followed the criteria described below, and were designed to increase reliability of manual annotation and reduce the number of errors generated in automatic tagging. This, in turn, facilitates more reliable annotation of syntactic information.

The English sections of the CHILDES database have been completely tagged with part-of-speech information using the CHILDES tag set (MacWhinney, 2000), following the additional criteria described in this section. This was done by using first the MOR program (MacWhinney, 2000) to generate ambiguous tags, then the POST program (Parisse & Le Normand, 2000) to disambiguate all the tags. The resultant corpora and tags have all been verified for adherence to the TalkBank XML schema (http://talkbank.org/talkbank.xsd).

The morphological annotation scheme

The English morphological analyzer incorporated in CHILDES produces various POS tags (the tag set contains thirty distinct tags), including ADJ (adjective), ADV (adverb), CO (communicator), CONJ (conjunction), DET (determiner), FIL (filler), N (noun), NUM (numeral), ON (onomatopoeia), PREP (preposition), PRO (pronoun), QN (quantifier), REL (relativizer) and V (verb). In most cases, the correct annotation of a word is obvious from the context in which the word occurs, but sometimes a more subtle distinction must be made. We discuss the two most common problematic issues below.

Adverb vs. preposition. Specific instances of certain words, such as *about, across, after, away, back, down, in, off, on, out, over* and *up*, belong to one of two categories: ADV and PREP. Previous versions of the CHILDES part-of-speech tag annotation for English also included a PTL (particle) tag to distinguish between adverbs and particles of phrasal verbs (e.g. *look **up** a phone number*). Because this distinction is often difficult to make, it was a source of frequent tagging errors and annotator disagreement. As a result, the PTL tag was retired, and in the current CHILDES annotation for English, the ADV tag is used for both adverbs and verbal particles. However, there is still a need to distinguish between cases when words should be tagged as either PREP or ADV. To correctly annotate them in context, we apply the following criteria.

First, a preposition must have a prepositional object, which is typically realized as a noun phrase. In some cases this noun phrase can be transformed, or even elided, but it is always possible to restore it when the word

709

in question is a preposition. Thus, in (1), *over* is a preposition, whereas in (2), it is not.

(1)   Somewhere *over* the rainbow.
(2)   The dog rolled *over*.

A preposition forms a constituent with its prepositional object, and hence is more closely bound to its object than an adverb or a particle would be to a noun phrase that happens to follow it. For example, in (3), *down* is not a preposition because *down the plate* does not form a constituent.

(3)   Put *down* the plate.

Adverbs and verbal particles, on the other hand, are only linked to a verb. As discussed, because of the difficulty in distinguishing between verbal particles and other adverbs linked to the verb, our annotation conflates both cases under the ADV category. Examples of words tagged as ADV (with verbs shown as context) include: *stand **up***, *lay **down***, *step **back***, *call **up***, *put **down***, *hold **on***, *go **away***, *jump **up***.

Prepositional objects can be fronted, whereas the noun phrases that happen to follow adverbs cannot. For example, from sentence (4) we can construct the noun phrase (5), which indicates that *on* is a preposition. In contrast, we cannot form the noun phrase (7) from (6), so we conclude that in this case *on* is not a preposition.

(4)   He sat *on* the chair.
(5)   The chair *on* which he sat.
(6)   The teacher picked *on* the student.
(7)   *The student *on* which the teacher picked.

Similarly, from sentence (8) we can construct the noun phrase (9), hence *up* is a preposition in (8). However, from (10) we cannot construct (11), hence *up* is not a preposition here.

(8)   She climbed *up* the chimney.
(9)   The chimney *up* which she climbed.
(10)   She filled *up* the bottle.
(11)   *The bottle *up* which she filled.

In addition, the two locative constructions *out_of* and *next_to* are treated as single prepositions.

Verbs vs. auxiliaries. Distinguishing between V and AUX can be difficult for the verbs *be*, *do* and *have*. The following tests are applied. First, if the target word is accompanied by a non-finite verb in the same clause, it is an auxiliary. In examples (12) and (13), *have* and *do* are auxiliaries.

(12)   I *have* had enough.
(13)   I *do* not like eggs.

710

A second test that we apply in such cases is fronting: in interrogative sentences, the auxiliary is moved to the beginning of the clause, as seen in examples (14) and (15). Main verbs, in contrast, do not move: the word *do* in example (16) is not an auxiliary, since (17) is ungrammatical.

(14)   *Have* I had enough?
(15)   *Do* I like eggs?
(16)   I *do* my homework religiously.
(17)   **Do* I my homework religiously?

However, these tests do not always work for the verb *be*, which may head a non-verbal predicate, as in (18) vs. (19). This is even more problematic when the predicate includes a form which is ambiguous as to whether it is verbal or not, as in (20). We adopt the following strategy: in verb–participle constructions headed by the verb *be*, if the participle is in the progressive aspect, *be* is labeled as an auxiliary. In (19), *be* would therefore be labeled as an auxiliary. Moreover, if the participle can be a main verb, as in (20), we label *be* as an auxiliary.

(18)   John *is* a teacher.
(19)   John *is* smiling.
(20)   John *is* finished.

The verb *have* can also be problematic. If the sentence is a verb–participle construction, we label *have* as an auxiliary. For example, in (21), *has* is labeled as an auxiliary. However, *have* in verb–infinitival constructions is labeled as a main verb: in (22), *have* and *drink* are both main verbs in separate clauses.

(21)   John *has* gone away.
(22)   John *has* to drink milk.

The syntactic annotation scheme

We represent syntactic information in CHILDES data in terms of LABELED DEPENDENCIES that correspond to GRAMMATICAL RELATIONS (GRs), such as subjects, objects and adjuncts. As in many flavors of dependency-based syntax (Mel'čuk, 1988; Hudson, 1984), each GR in our scheme represents a relationship between two words: a HEAD (also often referred to as a PARENT, REGENT or GOVERNOR), and a DEPENDENT (also often referred to as a CHILD or MODIFIER). In addition to the head and dependent words, a GR also includes a LABEL (or GR type) that indicates what kind of syntactic relationship holds between the two words. Figure 1 shows how the sentence *We eat the cheese sandwich* is annotated in our scheme. In this example, each arrow corresponds to a grammatical relation, where the head word points to the dependent word. For example, the arrow labeled SUBJ pointing from
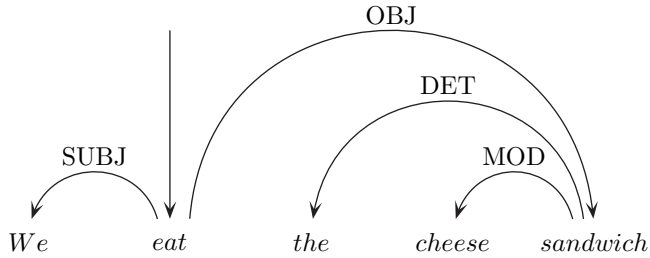
711

Fig. 1. Syntactic dependency structure for the sentence *We eat the cheese sandwich.*

eat to *we* indicates that *we* is the subject of *eat*. We can represent the same syntactic structure by simply listing each GR. Using the format LABEL(*head*, *dependent*), the list of GRs is: SUBJ(*eat*, *we*), OBJ(*eat*, *sandwich*), DET(*sandwich*, *the*), MOD(*sandwich*, *cheese*).

In choosing to represent syntactic information as grammatical relations, we considered issues relating to the use of the annotations, as well as to how the annotations can be produced efficiently and reliably. Grammatical relations are both intuitive to understand and straightforward to represent, which make them easier to annotate manually than representations that rely more heavily on embedded structures, such as phrase structure or constituent trees. Focusing on word-to-word relationships also makes the annotations easy to use, since each type of relation is meaningful both within a sentence and in isolation. For example, finding instances of pronouns used as subjects or direct objects is a simple matter of searching for SUBJ or OBJ annotations with the pronoun part-of-speech tag. In a representation based on constituents, however, a more complex search for a tree fragment involving noun phrases (NPs) and verb phrases (VPs) would be necessary. At the same time, representing GRs using labeled dependency trees allows for representation of much of the same hierarchical syntactic structure present in constituent structures, although some information about constituent boundaries is not represented. In addition, dependency structures are well-suited for high-accuracy parsing using data-driven natural language processing approaches (Buchholz & Marsi, 2006). This makes large-scale automatic annotation possible, starting only with a limited amount of manually annotated data. Finally, representation of syntax through dependency-based GRs allows us to focus on specific syntactic phenomena without many of the theory-specific assumptions necessary for syntactic representation with more complex formalisms. In choosing the specific grammatical relations to annotate in CHILDES, we made no attempt to push forward any particular theory of syntax, but simply to provide the information with which different theories and hypotheses can

be tested and validated empirically. That being said, our annotation does follow a set of guidelines, which we present in this section.

To define more specific characteristics of the syntactic structures in our scheme, it is helpful to think of the set of grammatical relations in an utterance as a dependency graph, as follows. We start by making each word in the utterance a node in a graph. Since GRs are represented as dependencies that hold between two words, we can define directed edges in the graph between each pair of words in a GR, pointing from the head to the dependent, as seen in Figure 1. We label each edge with the appropriate GR label. Finally, we add one extra node in the graph: a source node with no incoming edges. We then create edges from this source node to any previously existing nodes with no incoming edges (usually the main verb). We use the label ROOT for these edges. We consider a syntactic structure well-formed if the following conditions apply to the dependency graph:

- Each node that corresponds to a word has exactly one incoming edge.
- There is exactly one source node, with exactly one outgoing edge.
- There are no loops.

According to these conditions, every word in the sentence is a dependent in exactly one grammatical relation, and exactly one of these grammatical relations is the ROOT relation where a word is a dependent, and the head is not a word in the sentence. This means that, if we ignore the directionality of the edges, the dependency graph is a tree with labeled branches. If we ignore the extra node in the graph, the dependent of the ROOT relation is in fact the root of the dependency tree (in Figure 1, *eat* is the root).

Dependency-based grammatical relations. Our dependency-based annotation scheme distinguishes among several types of grammatical relations by specifying one of thirty-seven GR types as a label for each dependency (a head-dependent pair). The scheme includes labels for general GRs, such as subject, object and adjunct, and also finer distinctions within a general GR type (for example, whether an adjunct is a finite clause or a non-finite clause). In general, when a dependency exists between a function word (such as a determiner or a complementizer) and a content word (such as a noun or a verb), we make the content word the head, and the function word the dependent. For example, in the noun phrase *the boy*, the word *boy* is the head of *the*, and not the other way around. One notable exception to this general heuristic is prepositional phrases, where we make the preposition the head of its prepositional object. We describe the GRs in our scheme below.

Subjects are represented using one of four labels. The SUBJ label is used for simple subjects (subjects that are not clauses; typically noun phrases), ESUBJ for expletive subjects (semantically empty subjects, such as *there* in *there is/are* NP constructions), CSUBJ for finite clauses that fulfill a subject
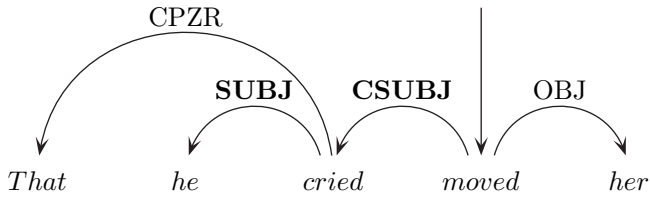
Fig. 2. Syntactic dependency structure for the sentence *That he cried moved her*.
Subject labels are shown in boldface.

relation as a dependent of a verb, and XSUBJ for non-finite clauses that fulfill a subject relation as a dependent of a verb (in general, prefixing a GR label in the annotation scheme with C or X indicates the dependent is the head of a finite or non-finite clause, respectively). Sentences (23–26) show examples of these different types of subjects. In each of these examples, the word or phrase in question is a subject of the verb that follows it. Note that in (25), *he* is a SUBJ of *cried*, while *that he cried* is a CSUBJ of *moved*. To represent this CSUBJ relation as a dependency, we make *moved* the head and *cried* the dependent. This is because *cried* is the root of the subtree formed by the dependencies in the phrase *that he cried* (it is the word that is not a dependent of another word within that phrase). Figure 2 shows a graphical representation of the dependency structure containing the GRs for sentence (25). These include SUBJ and CSUBJ, described above, in addition to other GRs introduced later in this section: OBJ (direct object) and CPZR (complementizer).

(23)   [*John*]$_{\text{SUBJ}}$ likes Mary.
(24)   [*There*]$_{\text{ESUBJ}}$ is a cup on the table.
(25)   [*That he cried*]$_{\text{CSUBJ}}$ moved her.
(26)   [*Eating vegetables*]$_{\text{XSUBJ}}$ is important.

Objects are represented by the labels OBJ (direct object), OBJ2 (the second object of a ditransitive verb, non-prepositional) and IOBJ (indirect object, a required verb complement introduced by a preposition). Verb complements that are realized as clauses are labeled COMP if they are finite and XCOMP otherwise, as seen in (27) and (28). Additionally, we mark required locative adjectival or prepositional phrase arguments of verbs as LOC (locative), as in (29) and (30). LOC is used instead of JCT, the adjunct label, when the prepositional phrase or adverb is required by the verb. This is especially relevant for words such as *here*, *there* and *back*, which would normally be labeled JCT.

(27)   I think [*that was Fraser*]$_{\text{COMP}}$.
(28)   You stop [*throwing the blocks*]$_{\text{XCOMP}}$.

(29) Put the toys [*in the box*]$_{\text{LOC}}$.
(30) Put the toys [*back*]$_{\text{LOC}}$.

Nominal, adjectival or prepositional phrase complement relations of copula (including not just *be*, but also verbs such as *become*, *seem* and *get*) are represented with the label PRED. For clausal predicates, as in (31) and (32), we use CPRED and XPRED.

(31) This is [*how I drink my coffee*]$_{\text{CPRED}}$.
(32) My goal is [*to win the contest*]$_{\text{XPRED}}$.

The label JCT is used for adjuncts: optional modifiers of verbs, adjectives or adverbs. CJCT and XJCT are used for finite and non-finite clausal adjuncts (sometimes referred to as 'adverbial subordinate clauses'). Sentences (33–36) illustrate the use of adjunct GR labels.

(33) That's [*much*]$_{\text{JCT}}$ better.
(34) Sit [*on the stool*]$_{\text{JCT}}$.
(35) You'll wear mittens [*when it snows*]$_{\text{CJCT}}$.
(36) He left [*after eating lunch*]$_{\text{XJCT}}$.

The labels MOD, CMOD and XMOD are used for modifiers or complements of nouns, as seen in (37–39).

(37) That's a [*nice*]$_{\text{MOD}}$ box.
(38) The boy [*who read the book*]$_{\text{CMOD}}$ is smart.
(39) Bigger feet [*to stand on*]$_{\text{XMOD}}$.

The annotation scheme also includes several GRs that are more easily identifiable: AUX (auxiliary verb or modal, typically dependents of verbs), NEG (negation), DET (determiner), QUANT (quantifier), POBJ (object of a preposition), CPZR (complementizer), COM (communicator), INF (infinitival *to*), VOC (vocative) and TAG (tag question). These are illustrated in sentences (40–46).

(40) Fraser [*is*]$_{\text{AUX}}$ [*not*]$_{\text{NEG}}$ reading [*a*]$_{\text{DET}}$ book.
(41) [*Some*]$_{\text{QUANT}}$ juice.
(42) On [*the stool*]$_{\text{POBJ}}$.
(43) Wait [*until*]$_{\text{CPZR}}$ the noodles are cold.
(44) [*Oh*]$_{\text{COM}}$, I took it.
(45) Thank you, [*Eve*]$_{\text{VOC}}$.
(46) You know how [*to*]$_{\text{INF}}$ count, [*don't you*]$_{\text{TAG}}$?

A special GR label, COORD, is used for representing coordination using dependencies. In a coordination structure, the head is a conjunction (usually *and*), and several types of dependents are possible. Once the COORD relations are formed between the head coordinator and each coordinated item
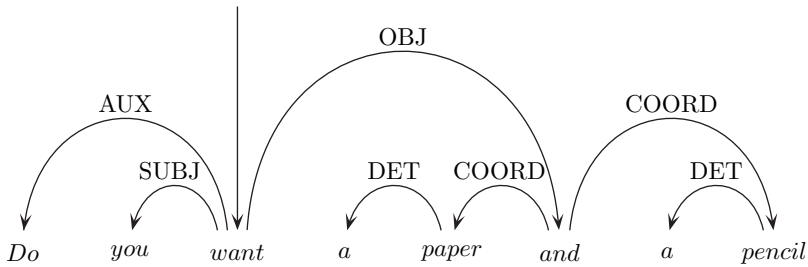
715

Fig. 3. Syntactic dependency structure for the sentence *Do you want a paper and a pencil?*

(as dependents), the coordinated phrase can be thought of as a unit represented by the head coordinator. For example, consider two coordinated verb phrases with a single subject (as in *I walk and run*), where two verbs are dependents in COORD relations to a head coordinator. The head of COORD is then also the head of a SUBJ relation where the subject is the dependent. This indicates that both verbs have that same subject. In the case of a coordinated string with multiple coordinators, the COORD relation applies compositionally from left to right. In coordinated lists with more than two items, but only one coordinator, the head coordinator takes each of the coordinated items as dependents. In the absence of an overt coordinator, the right-most coordinated item acts as the coordinator (the head of the COORD relation). Figure 3 shows the graphical representation of a coordinated structure. We note that the representation that the coordinated noun phrase *a paper and a pencil* is the object of *wanted* is accomplished by creating an OBJ relation with *and* as the dependent of *want*, and COORD relations for both *paper* and *pencil* as dependents of *and*.

Finally, we added some specific relations for handling problematic issues we encountered during data annotation. We use ENUM (enumeration) for constructions such as *one, two, three, go* or *a, b, c*. We use TOP (topicalization) to indicate an argument that is topicalized, as in (47). We use SRL (serial verb) in cases like (48) and (49). Finally, we mark sequences of proper names that form the same entity (e.g. *New York*) as NAME, and a date with month and year (e.g. *July 1980*), month and day (e.g. *July 10*) or month, day and year (e.g. *July 10, 1980*) as DATE. We arbitrarily set the month as the head, and the adjacent day and and/or year as its dependents.

(47)   [*Tapioca*]$_{\text{TOP}}$, there is no tapioca.
(48)   [*Come*]$_{\text{SRL}}$ see if we can find it.
(49)   [*Go*]$_{\text{SRL}}$ play with your toys.

Elision relations. Dependency structures represent the relationships between WORDS; when words (or entire phrases) are elided, binary relations

716

can be difficult to represent when one of the participants in the relation is not overtly present. To overcome this problem, we define a mechanism to indicate the presence of elided elements in an utterance. This is done by modifying the labels of GRs that involve elided elements. The bulk of these cases are true ellipsis, as in *Yes, I want to, too*, but the same mechanism is also used for partial utterances and interrupted utterances (such as *he has a …*).

Examples of how elision relations are represented through GR labels include AUX-ROOT (e.g. *Yes, I can*, where instead of assigning the ROOT label to the relation between *can* and the root node of the dependency tree, the prefix AUX- is added to the GR label, indicating that *can* would be the dependent of an AUX relation where the item that would serve as the head is elided); AUX-COMP (*I wish you would*, a similar case to the previous ONE, but here *would* is a dependent of *wish*, and the AUX- prefix indicates that elided material would be the head of an AUX relation with *would*, and the dependent of a COMP relation with *wish*); AUX-COORD (*and he will*); DET-OBJ, usually resulting from an interrupted utterance (as in *he has a*); DET-POBJ, which identifies a determiner of a noun with an elided prepositional object (*climb up the*); DET-COORD (*and a*); INF-XCOMP (*Yes, I want to, too*); INF-XMOD (*time to*); QUANT-OBJ (*you've just had some*); etc. These further emphasize the specific challenges of the corpus we deal with.

Format of the display. The format of the grammatical relation (GR) annotation, which we use in the examples that follow, associates with each word in the surface string a triple $i|j|g$, where i is the index of the word that serves as the dependent, j is the index of the dependent's syntactic head and g is the label corresponding to the grammatical relation represented by the syntactic dependency between the ith and j^{th} words. If the topmost head of the utterance is the ith word, it is labeled i|0|ROOT. For example, in:

```
the      tall     boy
1|3|DET  2|3|MOD  3|0|ROOT
```

the first word *a* is a DET of word 3 (*boy*), which is itself the ROOT of the utterance.

Manual annotation of the corpus

We focused our manual annotation on a set of CHILDES transcripts for a particular child, Eve (Brown, 1973), and we refer to these transcripts, distributed in a set of twenty files, as the EVE CORPUS. We manually annotated (including correction of POS tags) the first fifteen files of the Eve corpus following the GR scheme outlined above. The annotation process started

717

with the purely manual annotation of 5,000 words (approximately one Eve file). This initial annotated corpus was used to train a data-driven parser, as described in the next section. This parser was then used to label five additional Eve files automatically, followed by a thorough manual checking stage, where each syntactic annotation was manually verified and corrected if necessary. We retrained the parser with the newly annotated data, and proceeded in this fashion until fifteen files had been annotated and thoroughly manually checked.

Annotating child language proved to be challenging, and as we progressed through the data, we noticed grammatical constructions that the initial GR annotation scheme defined by Sagae *et al.* (2004) could not handle adequately. For example, the original GR scheme did not differentiate between locative arguments and locative adjuncts, so we created a new GR label, LOC, to handle required verbal locative arguments such as *on* in *put it on the table*. *Put* licenses a prepositional argument, and the existing POS tag PREP and the GR label JCT could not capture this requirement.

In addition to adding new GR types, we also faced challenges with telegraphic child utterances lacking verbs or other content words. For instance, *Mommy telephone* could have one of several meanings: 'Mommy this is a telephone', 'Mommy I want the telephone', 'that is Mommy's telephone', etc. (Bloom, 1970). We tried to be as consistent as possible in annotating such utterances and determined their GRs from context. It was often possible from preceding utterances to determine the VOC reading vs. the MOD ('Mommy's telephone') reading. If it was not possible to determine the correct annotation from context, we annotated such utterances as VOC relations. It is worth noting that, in such contexts, transcribers could make use of commas to mark the intonational drop that provides a good clue to the use of the vocative.

After annotating the fifteen Eve files, we had 18,843 fully manually corrected utterances (10,280 adult utterances and 8,563 child utterances). The utterances consist of 65,363 words, with one GR (labeled head-dependent pair) per word.

PARSING

Although the CHILDES annotation scheme proposed by Sagae *et al.* (2004) has been used in practice for automatic parsing of child language transcripts (Sagae *et al.*, 2004; 2005), such work relied mainly on a statistical parser (Charniak, 2000) trained on (and tuned for) *Wall Street Journal* articles, since a large enough corpus of annotated CHILDES data was not available. Having a corpus of 65,000 words of CHILDES data annotated with grammatical relations represented as labeled dependencies allows us to develop a parser tailored for the CHILDES domain.

718

Our overall parsing approach uses a best-first probabilistic shift-reduce algorithm, working left-to-right to find labeled dependencies one at a time. The algorithm is essentially a dependency version of the data-driven constituent parsing algorithm for LR-like probabilistic parsing described by Sagae & Lavie (2006). Because CHILDES syntactic annotations are represented as labeled dependencies, using a dependency parsing approach allows us to work with that representation directly.

We briefly describe the parsing approach below. A detailed description of the algorithm and learning strategy is included in Sagae & Tsujii (2007).

Shift-reduce dependency parsing

Left-to-right shift-reduce approaches have been shown to be capable of high levels of accuracy in dependency parsing (Nivre, Hall, Nilsson, Eryigit & Marinov, 2006). Here, we present an adaptation of the deterministic constituent parsing algorithm used by Sagae & Lavie (2006) to dependency parsing. The resulting algorithm is similar to the deterministic algorithm proposed by Nivre (2003), but follows a basic shift-reduce strategy similar to the one pursued by the classic LR parser (Knuth, 1965). This difference allows a probabilistic version of the algorithm to be seen as a natural extension of the GLR algorithm (Tomita, 1991) and its corresponding LR probabilistic model (Briscoe & Carroll, 1993).

The two main data structures in the algorithm are a stack S and a queue Q. S holds subtrees of the final dependency tree for an input sentence, and Q holds the words in an input sentence. S is initialized to be empty, and Q is initialized to hold every word in the input in order, so that the first word in the input is in the front of the queue.

The parser can perform two main types of actions: SHIFT and REDUCE. When a shift action is taken, a word is shifted from the front of Q, and placed on the top of S (as a tree containing only one node, the word itself). When a reduce action is taken, the two top items in S (s_1 and s_2) are popped, and a new item is pushed onto S. This new item is a tree formed by making the root (the only node with no parent in the tree) of s_1 a dependent of the root of s_2, or the root of s_2 a dependent of the root of s_1. Depending on which of these two cases occur, we call the action REDUCE-LEFT or REDUCE-RIGHT, according to whether the head of the new tree is to the left or to the right its new dependent. In addition to deciding on the direction of a reduce action, the parser must also choose the label of the newly formed dependency arc.

In order to choose the appropriate parser actions that result in a correct syntactic analysis for specific input, the parser uses maximum entropy models for classification (Berger, Della Pietra & Della Pietra, 1996). Maximum entropy classification is a machine-learning technique that aims to learn a mapping between classes and sets of features. In our case, the

classes are parser actions, and the features are derived from specific parser states (contents of the stack and input queue) where specific actions should be applied.

Although the algorithm is deterministic as described so far, the maximum entropy classifier assigns not just one action per set of features, but probabilities associated with several possible actions. This allows the parser to perform a search of the space of possible derivations to determine the analysis with highest probability. For further details, see Sagae & Tsujii (2007).

EVALUATION

To examine the reliability of grammatical relation annotations produced automatically by our parser, we evaluated its accuracy against manually corrected grammatical relations for two sets of transcripts in CHILDES, as described below.

Methodology

To evaluate the parser, we tested its output against the gold-standard files. We performed a fifteen-fold cross-validation evaluation on the fifteen manually curated files (to evaluate the parser on each file, the remaining fourteen files are used to train the parser). In addition to overall accuracy, we report the parser's performance on adult and child utterances separately. Finally, to evaluate the parser's portability to other CHILDES corpora, we also tested the parser on one file (516 utterances, approximately 1,600 words) from a different corpus in the CHILDES database, the Seth corpus.

The parser is highly efficient: training on the entire Eve corpus takes less than twenty minutes on a Linux workstation with a Pentium 4 1.8 GHz processor and 2 GB of RAM. Once the parser is trained, parsing the Eve corpus takes twenty seconds (over 4,000 GRs per second).

We report two evaluation measures: LABELED and UNLABELED dependency error rates. A labeled dependency is considered exactly correct if the dependency label and headword index produced by the parser match the gold-standard annotation (that is, the GR annotation produced for that head–dependent pair is perfect). For an unlabeled dependency to be correct only the headword index produced by the parser is required to match the gold-standard annotation, and the dependency labels are ignored (that is, the GR annotation produced contains the correct head–dependent pair, and the GR label may or may not be correct). These are the standard evaluation measures in the computational linguistics community for this kind of task (Buchholz & Marsi, 2006).

```
              go                buy             an              apple
parser:    1│2│COORD        2│0│ROOT      3│4│DET         4│1│OBJ
gold:      1│2│SRL          2│0│ROOT      3│4│DET         4│2│OBJ
```

Fig. 4. Example of parser error and corresponding gold-standard annotation.

For example, compare the parser output for the utterance *go buy an apple* to the gold standard (Figure 4). This sequence of GRs has two labeled dependency errors and one unlabeled dependency error. 1|2|COORD for the parser versus 1|2|SRL is a labeled error because the dependency label produced by the parser (COORD) does not match the gold-standard annotation (SRL), although the unlabeled dependency is correct, since the headword assignment, 1|2, is the same for both. On the other hand, 4|1|OBJ versus 4|2|OBJ is both a labeled dependency error and an unlabeled dependency error, since the headword assignment produced by the parser does not match the gold standard.

Results

By comparing the parser output on held-out data to the corresponding gold-standard annotations, we determine the number of labeled and unlabeled errors for each file. The error rates are calculated by dividing the total number of errors by the total number of grammatical relations. This can be done for the entire corpus, or for individual files.

The unlabeled dependency error rate is 4·71% and the labeled error rate is 6·09%. This means that the parser assigns the correct analysis for about 94% of the grammatical relations in the corpus. Performance in individual files ranged between the best unlabeled error rate of 3·42% and labeled error rate of 4·53% for the fifth file, and the worst error rates of 6·74% and 8·15% for unlabeled and labeled respectively in the fifteenth file.

To provide a more detailed view of the accuracy of the parser in the analysis of specific grammatical relations, Table 1 shows how the parser performs with respect to four common GRs according to the additional metrics of precision, recall and f-score, which are commonly used for parser evaluation in the computational linguistics literature. Precision for a specific GR is defined as the number of correct instances of that GR produced by the parser, divided by the total number of instances of that GR produced by the parser. In other words, precision measures how often the parser is correct when it outputs a specific GR. Conversely, recall for a specific GR is defined as the number of correct instances of that GR produced by the parser, divided by the total number of instances of that GR in the gold-standard annotation. For example, a system can achieve perfect recall (1·0) of SUBJ relations by producing a SUBJ relation for every possible

TABLE 1. *Precision, recall and f-score of four specific GRs*

GR	# occurrences	# errors	Precision	Recall	F-score
SUBJ	8651	303	0·958	0·955	0·956
OBJ	5669	324	0·941	0·930	0·935
XCOMP	801	146	0·867	0·812	0·839
LOC	514	40	0·950	0·923	0·936

head–dependent pair of words. This strategy, however, would result in very poor precision, since only a small fraction of those SUBJ relations would match the gold standard. Since it is possible in general to trade precision for recall, and vice versa, we also calculate the f-score, which is the harmonic mean of precision and recall. As an example, precision, recall and f-score are defined as follows for the SUBJ relation (where $SUBJ_{test}$ is the set of all SUBJ GRs in the parser's output, and $SUBJ_{gold}$ is the set of all SUBJ GRs in the gold-standard annotated corpus):

$$precision_{SUBJ} = \frac{\left|\{SUBJ_{test}\} \cap \{SUBJ_{gold}\}\right|}{\left|\{SUBJ_{test}\}\right|}$$

$$recall_{SUBJ} = \frac{\left|\{SUBJ_{test}\} \cap \{SUBJ_{gold}\}\right|}{\left|\{SUBJ_{gold}\}\right|}$$

$$F_{SUBJ} = \frac{2 \cdot precision_{SUBJ} \cdot recall_{SUBJ}}{precision_{SUBJ} + recall_{SUBJ}}$$

As Table 1 shows, the parser has high precision and recall for non-clausal verb complements (SUBJ, OBJ and LOC). For clausal verb complements, such as XCOMP, precision and recall are lower (although still reasonably high, well above 0·8). This is expected, since clausal complements are more syntactically complex, and therefore more difficult to identify, than non-clausal complements.

Testing child and adult utterances separately

We also tested accuracy of the parser on child utterances and adult utterances separately. To do this, we split the gold-standard files into child and adult utterances, producing gold-standard files for each. We then repeated the cross validation procedure described above (training the parser on every set of fourteen of the fifteen Eve files, and testing on the remaining file) for the child utterance files and for the adult utterance files.

Although the parser had high accuracy on both child and adult data, its accuracy was slightly better on the adult utterances. This might be due in part to greater grammatical uniformity (the complexity of the child

722

TABLE 2. *Error rates for child utterances and adult utterances in the Eve corpus*

	Unlabeled	Labeled
Child	5·66	7·21
Adult	4·18	5·44

TABLE 3. *Testing the parser on the Seth corpus, after training on the Eve corpus*

	Unlabeled	Labeled
Errors (total)	9·65	14·24
Errors (child)	15·07	23·25
Errors (adult)	5·44	7·22

utterances varies greatly, according to the child's level of language development), and to the larger size of the corpus containing training data for adult utterances. The results are shown in Table 2.

Testing on a different corpus

Because several aspects of the language in CHILDES transcripts vary from child to child, and because CHILDES transcripts may vary with respect to transcription standards, our parsing accuracy figures should be regarded as what the parser is capable of under ideal circumstances: training the parser on transcripts for a particular child, and testing on held-out data for the same child. In our final evaluation, we test the parser under highly adverse conditions, using test data that differs greatly from the material from the Eve corpus that was used for training. In this evaluation, our test set is one file from the Seth corpus (Wilson & Peters, 1988).

As expected, parser accuracy on the Seth corpus is not as high as it is on the Eve corpus, although it is still at reasonable levels (Table 3). The drop in accuracy is largely due to the poor performance on the child data, as the performance on the adult data comes close to that of tests on Eve. Although the error rates seem high for child utterances, this is due in great part to differences in the annotation of the Seth corpus, as explained in the analysis in the following section.

ERROR ANALYSIS

A major source for parser errors on the Eve corpus (112 out of 5181 errors) was 'telegraphic' speech, as in *Mommy telephone* or *Fraser tape + recorder*

723

floor. This type of speech may be the most challenging since, even for a human annotator, determining a GR is difficult. The parser usually labeled such utterances with the noun as the ROOT and the proper noun as the MOD, while the gold-standard annotation is context-dependent as described above. Possible ways to improve these results include more consistent manual annotation according to utterance-level criteria, and a parsing strategy that benefits from information about the context of utterances. The parser currently does not take context (neighboring utterances) into account, and therefore cannot learn to mimic manual annotation that is based explicitly on such context.

The XCOMP relation is also problematic, causing about 150 instances of parser error (these instances are at the word level; typically, the GRs for most words in these utterances are correct, despite the presence of an error). The majority of the errors in this category revolve around dropped words in the main clause, for example *want eat cookie*. Often, the parser labels such utterances with COMP GRs, because of the lack of *to*. Exclusive training on utterances of this type may resolve the issue. Many of the errors of this type occur with *want*: the parser could be conditioned to assign an XCOMP GR with *want* as the ROOT of an utterance.

The parser also has difficulty with the scope of coordination, which is the cause of 154 instances of parser error. For example, in *a birthday cake with Cathy and Becky*, the parser's analysis contains the incorrect coordination of the NPs [*a birthday cake with Cathy*]$_{NP}$ and [*Becky*]$_{NP}$, instead of the correct coordination of the NPs [*Cathy*]$_{NP}$ and [*Becky*]$_{NP}$, forming the object of the preposition *with*.

The performance drop on the Seth corpus can be explained by a number of factors. First and foremost, Seth is widely considered in the literature to be a child who is likely to invalidate any theory (Wilson & Peters, 1988). His speech exhibits false starts and filler syllables extensively, and his syntax violates many 'universal' principles. This is reflected in the annotation scheme: the Seth corpus, following the annotation of Peters (1983), is abundant with FILLER SYLLABLES (approximately 11% of all tokens). Because the parser was trained with material that did not contain any instances of filler syllables, it has no hope of analyzing those correctly. Because parser actions are determined according to context within the utterance, the presence of filler syllables often results in a cascade of errors. This problem is exacerbated in a left-to-right parsing approach by the fact that filler syllables usually occur near the beginning of an utterance (so the parser is guaranteed to make an early error, creating an unknown syntactic context for subsequent actions). On utterances in the Seth corpus that did not contain filler syllables, the accuracy of the parser was only slightly below its accuracy on Eve utterances. Another difficulty in the Seth corpus is the usage of DATES, of which there were no instances in the Eve corpus. The

parser had not been trained on the DATE GR and failed to parse it correctly.

The parser that was trained on the manually annotated Eve corpus was used to automatically annotate the entire English section of CHILDES. However, the CHILDES database includes transcripts in other languages that could also benefit from the same kind of morphosyntactic annotation. Recently, we have launched two projects that aim to adapt our automatic syntactic annotation techniques to other languages (from different language families). We already have promising results on the annotation of the SPANISH portion of CHILDES, and annotation of HEBREW is planned for the near future. We briefly discuss the Spanish annotation effort and our plans for the upcoming annotation of Hebrew, including some of the challenges we expect to face in adapting the annotation scheme.

Spanish

For Spanish syntactic annotation, we started by using the same annotation scheme as for English. Major relations, such as SUBJ or OBJ, are equally valid for both languages. However, some relations are expressed differently in the two languages, and automatic annotation may benefit from customization of the annotation scheme.

As an example, Spanish can insert a personal preposition *a* before an animate object. This was consistently annotated as the dependent in a JCT GR, and the following NP was therefore in a POBJ relation, as a dependent of the personal preposition. It could be advantageous to make an explicit distinction between a typical adjunct headed by a preposition, e.g. *de*, and one headed by the accusative *a*. In addition, we relied on the existing POS tags for the Spanish data, which were not always consistent with our syntactic scheme. For example, the verb *estar* was always tagged as a V (rather than AUX in certain contexts), and the parser therefore failed to assign it an AUX GR in several instances.

Such discrepancies can easily be solved with a more careful adaptation of the morphological analyzer and POS tagger used for Spanish, and minor revisions in the annotation scheme. Even in the face of these hurdles, the results of the Spanish parser are good. The Spanish corpus consisted of 10 files, totaling 12,854 utterances (roughly one-third child, two-thirds adult). Average utterance length, in words, is 3·26 (2·63 child, 3·56 adult). We manually annotated the entire corpus and evaluated by leave-one-out (training on nine files, then testing on the tenth, and averaging over ten such tests). The overall labeled error rate was 11·74%, and the overall unlabeled

725

TABLE 4. *Precision, recall and f-score for specific GRs in automatic annotation of Spanish transcripts*

GR	Precision	Recall	F-score
SUBJ	0·71	0·74	0·72
OBJ	0·78	0·83	0·80
JCT	0·73	0·76	0·75
XCOMP	0·86	0·91	0·89

error rate was 7·8%. This means that the parser was correct in over 88% of its GR assignments for the Spanish data. Unlike in the English evaluation, the set of 10 annotated files contained data for five different children, so these accuracy figures are less likely to be artificially high. Labeled error rates for individual files ranged from a low of 9·01% to a high of 14·22%.

Table 4 shows the precision, recall and f-score for four particular GRs. We see that the accuracy for SUBJ in Spanish is far below the accuracy for SUBJ in English. Surprisingly, however, we also see that the accuracy for XCOMP is higher than it is for XCOMP in the English data. This improvement is likely a reflection of how different features from the data can affect the parser's behavior. In English, the parser learned to associate infinitives (which often appear in XCOMP relations) with the infinitive particle *to*. In the test data, the particle *to* was often absent, causing the parser to favor other (incorrect) interpretations. In Spanish, however, infinitives (and other non-finite forms, which are also associated with XCOMP) are marked with suffixes, and therefore readily identifiable.

Upcoming work : Hebrew

Annotation of the Hebrew data will also require revision of the morphosyntactic annotation scheme. For example, Hebrew lacks independent particles, and encodes spatio-directional elements such as path and goal through monolexemic verbs (Berman, 1979). The morphological analyzer (and corresponding POS tagset) that was developed for Hebrew reflects the rich semantic information provided by the combination of tri-consonantal roots with different verb patterns (*binyanim*), and thus provides information concerning not only different lexical categories (nouns, verbs, adjectives, prepositions, pronouns, etc.), but also detailing inflectional and derivational paradigms, including tense, number, gender and person.

Several Hebrew-specific issues need particular treatment through the syntactic annotation scheme. For example, in copula constructions the copula may be explicitly manifested as some form of the verb *be* (in past and future tense); or as suppletive morphemes which are identical to third person nominative or demonstrative pronouns; or it may be altogether missing, in

726

the present tense (Berman, 1978). As another example, Hebrew has verbless predicates with no counterparts in English (Doron, 1983). Such constructions necessitate specifically tailored dependency relations.

The Hebrew corpus that will be used for training the syntactic parser includes phonemically transcribed child–caretaker interactions from two major sources: The Berman longitudinal corpus, with data from four children between the ages of 1;06 and 3;05; and the Ravid longitudinal corpus, consisting of data from two siblings between the ages of 0;09 to around 6;0. Together, the Hebrew corpus includes 305 files, totaling 112,258 utterances (47,360 child utterances), with a mean length of 3·76 words per utterance.

We are actively pursuing the tasks of annotating these data. To date, the morphological analyzer was successfully applied to some three-quarters of the child-directed utterances, and about one-quarter of the child speech. Due to the sophisticated transliteration of the data, the level of morphological ambiguity is very low (approximately 20% of the analyzed tokens are ambiguous). As soon as we complete the morphological analysis of the entire corpus, we will train a POS tagger to disambiguate the results, and will then move on to syntactic annotation.

CONCLUSION

We described an annotation scheme for representing syntactic information as grammatical relations in CHILDES data, a manually curated gold-standard corpus of 65,000 words annotated according to this GR scheme, and a parser that was trained on the annotated corpus and produces highly accurate grammatical relations for both child and adult utterances. The parser was used to automatically assign grammatical relations to the entire English section of CHILDES. These resources are now freely available to the research community, and we expect them to be instrumental in psycholinguistic investigations of language acquisition and child language development.

Our immediate plans include continued improvement of the parser, and the application of automatic morphosyntactic analysis to other languages. Dependency schemes based on functional relationships exist for a number of languages (Buchholz & Marsi, 2006), and the general parsing techniques used in the present work have been shown to be effective in several of them (Nivre et al., 2006). We have already obtained promising results on the application of the parser to Spanish, and its application to Hebrew is under way.

This work illustrates the use of current techniques in computational linguistics with machine learning to aid child language research. Possible extensions include the use of computational models for analysis of semantic

roles and discourse structure. Additionally, it is our hope that the availability of a large corpus of syntactic analyses will fuel further research on models of child language acquisition based on these data, and encourage the implementation of new and existing models that can be evaluated and validated on the naturally occurring distribution of language phenomena reflected in CHILDES transcripts.

REFERENCES

Berger, A., Della Pietra, S. A. & Della Pietra, V. J. (1996). A maximum entropy approach to natural language processing. *Computational Linguistics* **22**(1), 39–71.

Berman, R. A. (1978). *Modern Hebrew structure*. Tel Aviv: University Publishing Projects.

Berman, R. A. (1979). Lexical decomposition and lexical unity in the expression of derived verbal categories in modern Hebrew. *Afroasiatic Linguistics* **6**, 1–26.

Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: MIT Press.

Bod, R. (2009). From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science* **33**(5), 752–93.

Borensztajn, G., Zuidema, J. & Bod, R. (2009). Children's grammars grow more abstract with age – evidence from an automatic procedure for identifying the productive units of language. *Topics in Cognitive Science* **1**, 175–88.

Briscoe, T. & Carroll, J. (1993). Generalised probabilistic lr parsing of natural language (corpora) with unification-based grammars. *Computational Linguistics* **19**(1), 25–59.

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

Buchholz, S. & Marsi, E. (2006). Conll-x shared task on multilingual dependency parsing. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CONLL-x)*, 149–64. New York City: Association for Computational Linguistics.

Charniak, E. (2000). A maximum-entropy-inspired parser. In *Proceedings of the First Conference of the North American Chapter of the Association for Computational Linguistics*, 132–39. San Francisco, CA: Morgan Kaufmann Publishers Inc.

Doron, E. (1983). Verbless predicates in Hebrew. Unpublished doctoral dissertation, University of Texas at Austin.

Hudson, R. A. (1984). *Word grammar*. Oxford: Basil Blackwell.

Knuth, D. (1965). On the translation of languages from left to right. *Information and Control* **8**(6), 607–639.

Lee, L. (1974). *Developmental sentence analysis*. Evanston, IL: Northwestern University Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*, 3rd edn. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (2008). Enriching CHILDES for morphosyntactic analysis. In H. Behrens (ed.), *Corpora in language acquisition research: History, methods, perspectives*, Vol. 6, 165–98. Amsterdam: Benjamins.

Mel'čuk, I. A. (1988). *Dependency syntax: Theorie and practice*. Albany, NY: SUNY Press.

Nivre, J. (2003). An efficient algorithm for projective dependency parsing. In *Proceedings of the Eighth International Worskshop on Parsing Technologies (IWPT)*, 149–60. Nancy.

Nivre, J., Hall, J., Nilsson, J., Eryigit, G. & Marinov, S. (2006). Labeled pseudo-projective dependency parsing with support vector machines. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*, 221–25. New York: Association for Computational Linguistics.

Parisse, C. & Le Normand, M.-T. (2000). Automatic disambiguation of the morphosyntax in spoken language corpora. *Behavior Research Methods, Instruments and Computers* **32**, 468–81.

Peters, A. M. (1983). *The units of language acquisition*. New York: Cambridge University Press.

Sagae, K. & Lavie, A. (2006). A best-first probabilistic shift-reduce parser. In *Proceedings of the Coling/ACL Poster Session*, 691–98. Sydney: Association for Computational Linguistics.

Sagae, K., Lavie, A. & MacWhinney, B. (2004). Adding syntactic annotations to transcripts of parent–child dialogs. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation* (*LREC 2004*), 1815–18. Lisbon: European Language Resources Association.

Sagae, K., Lavie, A. & MacWhinney, B. (2005). Automatic measurement of syntactic development in child language. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics* (*ACL'05*), 197–204. Ann Arbor, MI: Association for Computational Linguistics.

Sagae, K. & Tsujii, J. (2007). Dependency parsing and domain adaptation with LR models and parser ensembles. In *Proceedings of the CONLL Shared Task Session of the Joint Conferences on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (*EMNLP-CONLL 2007*), 1044–50. Prague: Association for Computational Linguistics.

Scarborough, H. S. (1990). Index of productive syntax. *Applied Psycholinguistics* **11**, 1–22.

Tomita, M. (ed.) (1991). *Generalized LR parsing*. Boston: Kluwer Academic Publishing.

Wilson, B. & Peters, A. M. (1988). What are you cookin' on a hot?: A three-year-old blind child's 'violation' of universal constraints on constituent movement. *Language* **64**, 249–73.