Measuring Language Development in Early Childhood Education: A Case Study of Grammar Checking in Child Language Transcripts

Khairun-nisa Hassanali
Computer Science Department
The University of Texas at Dallas
Richardson, TX, USA
nisa@hlt.utdallas.edu

Yang Liu
Computer Science Department
The University of Texas at Dallas
Richardson, TX, USA
yangl@hlt.utdallas.edu

Abstract

Language sample analysis is an important technique used in measuring language development. At present, measures of grammatical complexity such as the Index of Productive Syntax (Scarborough, 1990) are used to measure language development in early childhood. Although these measures depict the overall competence in the usage of language, they do not provide for an analysis of the grammatical mistakes made by the child. In this paper, we explore the use of existing Natural Language Processing (NLP) techniques to provide an insight into the processing of child language transcripts and challenges in automatic grammar checking. We explore the automatic detection of 6 types of verb related grammatical errors. We compare rule based systems to statistical systems and investigate the use of different features. We found the statistical systems performed better than the rule based systems for most of the error categories.

1 Introduction

Automatic grammar checking and correction has been used extensively in several applications. One such application is in word processors where the user is notified of a potential ungrammatical sentence. This feature makes it easier for the users to detect and correct ungrammatical sentences. Automatic grammar checking can also be beneficial in language learning where students are given suggestions on potential grammatical errors (Lee and Seneff, 2006). Another application of grammar checking is in improving a parser's performance for ungrammatical sentences. Since most parsers are trained on written data consisting mostly of grammatical sentences, the parsers face issues when parsing ungrammatical sentences. Automatic detection and correction of these ungrammatical sentences would improve the parser's performance by detecting the ungrammatical sentences and performing a second parse on the corrected sentences (Caines and Buttery, 2010). From an education perspective, measuring language skills has been extensively explored. There are systems in place that automatically detect and correct errors for second language learners (Eeg-Olofsson and Knuttson, 2003; Leacock et al., 2010).

One method used in measuring language development is the analysis of transcripts of child language speech. Child language transcripts are samples of a child's utterances during a specified period of time. Educators and speech language pathologists use these samples to measure language development. In particular, speech language pathologists score these transcripts for grammatical measures of complexity amidst other measures. Since manual analysis of transcripts is time consuming, many of these grammatical complexity measures require the speech language pathologists to look for just a few examples. The Index of Productive Syntax (IPSyn) (Scarborough, 1990) is one such measure of morphological and syntactic structure developed for measuring language samples of preschool children. The advantage of measures such as IPSyn is that they give a single score that can be used to holistically measure language development. However, they focus on grammatical constructs that the

Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications, pages 87–95,
Portland, Oregon, 24 June 2011.

child uses correctly and do not take into account the number and type of grammatical errors that are made by the child.

Educators wishing to measure language development and competence in a child will benefit from having access to the grammatical errors made by a child. Analysis of these grammatical errors will enable educators and speech language pathologists to identify shortcomings in the child's language and recommend intervention techniques customized to the child. Since manual identification of grammatical errors is both cumbersome and time consuming, a tool that automatically does grammar checking would be of great use to clinicians. Additionally, we see several uses of automatic grammar detection. For example, we can use the statistics of grammatical errors as features in building classifiers that predict language impairment. Furthermore, we could also use the statistics of these grammatical errors to come up with a measure of language development that takes into account both grammatical competence and grammatical deficiencies.

In this paper, we use existing NLP techniques to automatically detect grammatical errors from child language transcripts. Since children with Language Impairment (LI) have a greater problem with correct usage of verbs compared to Typically Developing (TD) children (Rice et al., 1995), we focus mainly on verb related errors. We compare rule based systems to statistical systems and investigate the use of different features. We found the statistical systems performed better than the rule based systems for most error categories.

2 Related Work

While there has been considerable work (Sagae et al., 2007) done on annotating child language transcripts for grammatical relations, as far as we know, there has been no work done on automatic grammar checking of child language transcripts. Most of the existing work in automatic grammar checking has been done on written text. Spoken language on the other hand, presents challenges such as disfluencies and false restarts which are not present in written text. We believe that the specific research challenges that are encountered in detecting and correcting child language transcripts warrant a more detailed examination.

Caines and Buttery (2010) focused on identifying sentences with the missing auxiliary verb in the progressive aspect constructions. They used logistic regression to predict the presence of zero auxiliary occurrence in the spoken British National Corpus (BNC). An example of a zero auxiliary construction is "You talking to me?". They first identified constructions with the progressive aspect and annotated the constructions for the following features: subject person, subject case, perfect aspect, presence of negation and use of pronouns. Their model identified zero auxiliary constructions with 96.9% accuracy. They also demonstrated how their model can be integrated into existing parsing tools, thereby increasing the number of successful parses for zero auxiliary constructions by 30%.

Lee and Seneff (2008) described a system for verb error correction using template matching on parse trees in two ways. Their work focused on correcting the error types related to subject-verb agreement, auxiliary agreement and complementation. They considered the irregularities in parse trees caused by verb error forms and used n-gram counts to filter proposed corrections. They used the AQUAINT Corpus of English News Text to detect the irregularities in the parse trees caused by verb error forms. They reported an accuracy of 98.93% for verb errors related to subject-verb agreement, and 98.94% for verb errors related to auxiliary agreement and complementation. Bowden and Fox (2002) developed a system to detect and explain errors made by non-native English speakers. They used classification and pattern matching rules instead of thorough parsing. Their system searched for the verb-related errors and noun-related errors one by one in one sentence by narrowing down the classification of errors. Lee and Seneff (2006) developed a system to automatically correct grammatical errors related to articles, verbs, prepositions and nouns.

Leacock et al. (2010) discuss automated grammatical error detection for English language learners. They focus on errors that language learners find most difficult - constructions that contain prepositions, articles, and collocations. They discuss the existing systems in place for automated grammatical error detection and correction for these and other classes of errors in a number of languages. Addi-

Label	Meaning	Example
0	No error	I like it.
1	Missing auxiliary verb	You talking to me?
2	Missing copulae	She lovely.
3	Subject-auxiliary verb agreement	You is talking to me.
4	Incorrect auxiliary verb used e.g. using does instead of is	She does dead girl.
5	Missing verb	She her a book.
6	Wrong verb usage including subject-verb disagreement	He love dogs.
7	Missing preposition	The book is the table.
8	Missing article	She ate apple.
9	Missing subject before verb	I know loves me.
10	Missing infinitive marker "to"	I give it her.
11	Other errors not covered in 1-10	The put.

Table 1: Different types of errors considered in this study

tionally, they touch on error annotations and system evaluation for grammatical error detection.

3 Data

For the purpose of our experiments, we used the Paradise dataset (Paradise et al., 2005). This dataset contains 677 transcripts corresponding to 677 children aged six that were collected in the course of a study of the relationship of otitis media and child development. The only household language spoken by these children was English. The transcripts in the Paradise set consist of conversations between a child and his/her caretaker. We retained only the child's utterances and removed all other utterances. The Paradise dataset (considering only the child's utterances) contains a total of 108,711 utterances, 394,290 words, and an average Mean Length of Utterance of 3.64. Gabani (2009) used scores on the Peabody Picture Vocabulary Test (Dunn, 1965), total percentage phonemes repeated correctly on a non-word repetition task and mean length of utterance in morphemes to label these transcripts for language impairment. A transcript was labeled as having been produced by a child with LI if the child scored 1.5 or more standard deviations below the mean of the entire sample on at least two of the three tests. Of the 677 transcripts, 623 were labeled as TD and 54 as LI.

We manually annotated each utterance in the transcripts for 10 different types of errors. Table 1 gives the different types of errors we considered along with examples. We focused on these 10 different types of errors since children with LI have problems with the usage of verbs in particular. The list of errors we arrived at was based on the errors we observed in the transcripts. Since an utterance could have more than one error, we annotated each utterance in the transcript for all the errors present in the utterance. While annotating the utterances, we observed that there were utterances that could correspond to multiple types of error. For example, consider the following sentence: "She go to school". The error in this sentence could be an error of a missing auxiliary and a wrong verb form in which case the correct sentence would be "She is going to school"; or it could be a missing modal, in which case the correct form would be "She will go to school"; or it could just be a subject-verb disagreement in which case "She goes to school" would be the correct form. Therefore, although we know that the utterance definitely has an error, it is not always possible to assign a single error. We also observed several utterances had both a missing subject and a missing auxiliary verb error. For example, instead of saying "I am going to play", some children say "Going to play", which misses both the subject and auxiliary verb. In this case, the utterance was annotated as having two errors: missing subject and missing auxiliary. Finally, single word utterances were labeled as being correct.

Table 2 gives the distribution of the errors in the corpus and percentage of TD and LI population that

No	Error Type	Percentage (Count)	% of LI children making error	% of TD children making error
1	Missing auxiliary	8.43% (641)	7%	5%
2	Missing copulae	36.67% (2788)	77.78%	45%
3	Subject-auxiliary agreement	6.31% (480)	40.74%	35%
4	Incorrect auxiliary verb used	0.71%(54)	11.47%	3%
5	Missing verb	5% (380)	29.63%	10%
6	Wrong verb usage	14.59% (1109)	68.5%	50%
7	Missing preposition	5% (380)	7.4%	5%
8	Missing article	3.97% (302)	29.63%	35%
9	Missing subject	7.69% (585)	3.7%	5%
10	Missing infinitive marker "To"	1.58% (120)	7.5%	11.67%
11	Other errors	10.05% (764)	56.7%	23.2%

Table 2: Statistics of Errors

made the error at least once in the entire transcript. As we can see from Table 2, 36.67% of the errors in the corpus are due to missing copulae. Wrong verb usage was the next most common error contributing to 14.59% of the errors in the corpus. We observed that there was a higher percentage of children with LI that made errors on all error categories except for errors related to missing article and missing subject. We observed that on average, the transcripts belonging to children with LI had fewer utterances as compared to transcripts belonging to TD children. Additionally, children with LI used many single word and two word utterances.

One annotator labeled the entire corpus for grammatical errors. To calculate inter-annotator agreement, we randomly selected 386 utterances annotated by the first annotator with different error types. The second annotator was provided these utterances along with the labels given by the first annotator[1]. In case of a disagreement, the second annotator provided a different label/labels. The annotator agreement using the average Cohen's Kappa coeffiecient was 77.7%. Out of the 386 utterances, there were 43 disagreements between the annotators. We found that for some error categories such as the missing auxiliary, there was high inter-annotator agreement of 95.32%, whereas for other categories such as wrong verb usage and missing articles, there was less agreement (64.2% and 65.3% respectively). In particular, we found low inter-annotator agreement on utterances that have errors that could be assigned to multiple categories.

4 Experiments

The transcripts were parsed using the Charniak parser (Charniak, 2000). Since the Paradise dataset consists of children's utterances, and many of them have not mastered the language, we observed that processing these transcripts is challenging. As is prevalent in spoken language corpora, these transcripts had disfluencies, false restarts and incomplete utterances, which sometimes pose problems to the parser.

We conducted experiments in detecting errors related to the usage of the -ing participle, subject-auxiliary agreement, missing copulae, missing verb, subject-verb agreement and missing infinitive marker "to". For each of these categories, we constructed one rule based classifier using regular expressions based on the parse tree structure, an alternating decision tree classifier that used rules as features and a naive Bayes multinomial classifier that used a variety of features. For every category, we performed 10 fold cross validation using all the utterances. We used the naive Bayes multinomial classifier and the alternating decision tree classifier from the WEKA toolkit (Hall et al., 2009). Table 3 gives the results using the three classifiers for the different categories of errors, where (P/R) F1 stands for (Pre-

[1]We will perform independent annotation of the errors and calculate inter-annotator agreement based on these independent annotations

Error	Rule Based System (P/R)F1	Decision Tree Classifier using Rules as features (P/R)F1	Naive Bayes Classifier using a variety of features (P/R)F1
Usage of -ing participle	(0.984/0.978) 0.981	**(0.986/1) 0.993**	(0.736/0.929) 0.821
Missing copulae	(0.885/0.9) 0.892	**(0.912/0.94) 0.926**	(0.82/0.86) 0.84
Missing verb	(0.875/**0.932**) 0.903	(**0.92**/0.89) **0.905**	(0.87/0.91) 0.9
Subject-auxiliary agreement	(0.855/0.932) 0.888	(**0.95**/0.84) 0.892	(0.89/**0.934**) **0.912**
Subject-verb agreement	(0.883/**0.945**) 0.892	(**0.92**/0.877) 0.898	(0.91/0.914) **0.912**
Missing infinitive marker "To"	**(0.97/0.954) 0.962**	(0.94/0.84) 0.887	(0.95/0.88) 0.914
Overall	(0.935/0.923) 0.929	(0.945/0.965) 0.955	**(0.956/0.978) 0.967**

Table 3: Detection of errors using rule based system, alternating decision tree classifier and naive Bayes classifier

No	Feature	Type
1	Verb Adjective	Bigram
2	Auxiliary Noun	Bigram
3	Auxiliary Progressive-verb	Bigram
4	Pronoun Auxiliary	Bigram
5	Wh-Pronoun Progressive verb	Bigram
6	Progressive-verb Wh-adverb	Bigram
7	Adverb Auxiliary	Skip-1
8	Pronoun Auxiliary	Skip-1
9	Wh-adverb Progressive-verb	Skip-1
10	Auxiliary Preposition	Skip-2

Table 4: Top most bigram features useful for detecting misuse of -ing participle

cision/Recall) F1-measure. Below we describe the different experiments we conducted.

4.1 Misuse of the -ing Participle

The -ing participle can be used as a progressive aspect, a verb complementation, or a prepositional complementation. In the progressive aspect, it is necessary that the progressive verb be preceded by an auxiliary verb. When used as a verb complementation, the -ing participle should be preceded by a verb and similiarly when used as a prepositional complement, the -ing participle should be preceded by a preposition.

Rule based system

The -ing participle is denoted by the VBG tag in the Penn tree bank notation. VP and PP correspond to the verb phrase and prepositional phrase structures respectively. The rules that we formed were as follows:

1. Check that the utterance has a VBG tag (if it does not have a VBG tag, it does not contain an -ing participle).
2. If none of the following conditions are met, there is an error in the usage of -ing participle:
   (a) The root of the subtree that contains the -ing participle should be a VP with the head being a verb if used as a verb complementation
   (b) The root of the subtree that contains the -ing participle should be a PP if used as a prepositional complement
   (c) The root of the subtree that contains the -ing participle should be a VP with the head being an auxiliary verb if used as a progressive aspect

Predictive model

The features that we considered were:

1. Bigrams from POS tags
2. Skip bigrams from POS tags

   We used the skip bigrams to account for the fact that there could be other POS tags between an auxiliary verb and the progressive aspect of the verb such as adverbs. A skip-n bigram is a sequence of 2 POS tags with a distance of n between them. We used skip-1 and skip-2 bigrams in this study.

Analysis

As we can see from Table 3, the alternating decision tree classifier with rules as features gave the best results with an F1-measure of 0.993. Table 4 gives the topmost 10 features extracted using feature selection. We got the best results when we used the reduced set of features as opposed to using all bigrams and skip-1 and skip-2 bigrams. We also used the results reported by (Caines and Buttery, 2010) to see if their method was successful in identifying zero auxiliary constructs on our corpus. When we used logistic regression with the coefficients and features used by (Caines and Buttery, 2010), we got a recall of 0%. When we trained the logistic regression model on our data with their features, we got a precision of 1.09%, recall of 53.6% and F1-measure of 2.14%. This leads us to conclude that the features that were used by them are not suitable for child language transcripts. Additionally, we also observed that based on the features they used, in some cases it is difficult to distinguish zero auxiliary constructs from those with auxiliary constructs. For example, "You talking to me?" and "Are you talking to me?" would have the same values for their features, although the former is a zero auxiliary construct and the latter is not.

4.2 Identifying Missing Copulae

A copular verb is a verb that links a subject to its complement. In English, the most common copular verb is "be". Examples of sentences that contain a copular verb is "She is lovely" and "The child who fell sick was healthy earlier". An example of a sentence that misses a copular verb is "She lovely".

Rule based system

The rule that we used was as follows:

If an Adjective Phrase follows a noun phrase, or a Noun phrase follows a noun phrase, the likelihood that the utterance is missing a copular verb is quite high. However, there are exceptions to such rules, for example, "Apple Pie". We formed additional rules to identify such utterances and examined their parse trees to determine the function of the two noun phrases.

Predictive model

The features we used were as follows:

1. Does the utterance contain a noun phrase followed by a noun phrase?
2. Does the utterance contain a noun phrase followed by an adjective phrase?
3. Is the parent a verb phrase?
4. Is the parent a prepositional phrase?
5. Is the parent the root of the parse tree?
6. Is there an auxiliary verb or a verb between the noun phrase and/or adjective phrase?

Analysis

As we can see from Table 3, the alternating decision tree classifier performed the best with an F1-measure of 0.926. Our rules capture simple constructs that are used by young children. The majority of the utterances that missed a copulae consisted of noun phrase and an adjective phrase or a noun phrase and a noun phrase. Hence, the rules based system performed the best. Some of the false positives were due to utterances like "She an apple" where it is unlikely that the missing verb is a copular verb.

4.3 Identifying Missing Verbs

Errors of this type occur when a sentence is missing the verb. For example, the sentence "You can an apple" lacks the main verb after the modal verb "can". Similarly, "I did not it" lacks a main verb after "did not". For the purpose of this experiment, we consider only utterances that contain a modal or an auxiliary verb but do not have a main verb. We also consider utterances that use the verb "do" and detect the main missing verb in such cases.

Rule based system

The rule we used was to check if the utterance contains an auxiliary verb or a modal verb but not a main verb. In this case, the utterance is definitely missing a main verb. In order to identify utterances where the words "did", "do" and "does" are auxiliary verbs, we use the following procedure: If the negation "not" is present after *did/do/does*, then *did/do/does* is an auxiliary verb and needs to be followed by a main verb. In the case of the utterance being a question, the presence of *did/do/does* at the beginning of the utterances indicates the use as an auxiliary verb. In such a case, we need to check for the presence of a main verb. The same holds for the other auxiliary verbs.

Predictive model
We used the following as features:

1. Is an auxiliary verb present?
2. Is a modal verb present?
3. Is a main verb present after the auxiliary verb?
4. Is a main verb present after the modal verb?
5. Type of utterance - interrogative, declarative
6. Is a negation (not) present?

Analysis
As we can see from Table 3, the alternating decision tree classifier using rules as features gave the best result with an F1-measure of 0.905. At present, we handle only a subset of missing verbs and specifically those verbs that contain an auxiliary verb. Since most of the utterances are simple constructs, the alternating decision tree classifier performs well.

4.4 Identifying Subject-auxiliary Agreement

In the case of the subject-auxiliary agreement and subject-verb agreement, the first verb in the verb phrase has to agree with the subject unless the first verb is a modal verb. In the sentence "The girls has bought a nice car", since the subject "The girls" is a plural noun phrase, the auxiliary verb should be in the plural form. While considering the number and person of the subject, we take into account whether the subject is an indefinite pronoun or contains a conjunction since special rules apply to these cases. Indefinite pronouns are words which replace nouns without specifying the nouns they replace. Some indefinite pronouns such as *all*, *any* and *more* take both singular and plural forms. On the other hand, indefinite pronouns like *somebody* and *anyone* always take the singular form.

Rule based system
The rule we used to identify subject-auxiliary agreement was as follows:

1. Extract the number (singular, plural) of the subject and the auxiliary verb in the verb phrase.
2. If the number of the subject and auxiliary verb do not match, there is a subject-auxiliary agreement error.

Predictive model
The features were as follows:

1. Number of subject - singular or plural
2. Type of noun phrase - pronoun or other noun phrase
3. Person of noun phrase - first, second, third
4. Presence of a main verb in the utterance (we are looking at the agreement only for the auxiliary verb)

Analysis
As we can see from Table 3, the naive Bayes multinomial classifier performed the best with an F1-measure of 0.912. We found that our system did not detect the subject-auxiliary agreement correctly if there was an error in the subject such as number agreement.

4.5 Identifying Subject-verb Agreement

In order to achieve subject-verb agreement, the number and person of the subject and verb must agree. The subject-verb agreement applies to the first verb in the verb phrase. We consider cases wherein the first verb is a main verb or contains a modal verb. An example of a sentence that has subject-verb disagreement is "The boy have an ice cream". The number and person of the subject "The boy" and the verb "have" do not match.

Rule based system
The rule we used to identify subject-verb agreement was as follows:

1. Extract the number (singular, plural) and person (first, second, third) of the subject and the first verb in the verb phrase.
2. If the verb is not a modal verb and the number and person of the subject and verb do not match, there is a subject-verb agreement error.

Predictive model
We used the following features to be used in a statistical setup:

1. Type of sentence - interrogative or declarative
2. Number of subject - singular or plural
3. Person of subject if pronoun - first, second or third
4. Number of verb - singular or plural
5. Person of verb - first, second or third
6. Type of verb - modal, main

Analysis
We found that our system did not detect errors in cases where there was a number disagreement. For example, in the sentence "The two dog is playing", our system based on the POS tag would assume that the subject is singular and therefore there is no subject-verb error. One way to improve this would be to detect number disagreement in the subject and correct it before detecting the subject-verb agreement.

4.6 Identifying Missing Infinitive Marker "To"

Errors of this type occur when the sentence lacks the infinitive marker "to". An example of such a sentence would be "She loves sleep". In this case, "She loves to sleep" would be the correct form. On the other hand, this statement is ambiguous since *sleep* could be used as a noun sense or a verb sense. We concentrated on identifying utterances that have the progressive verb followed by the verb in the infinitive form. Examples of such sentences are: "She is going cry". In this case, we can see that the sentence is missing the "to".

Rule based system
If the utterance contains a progressive verb followed by a verb in its infinitive form, it is missing the infinitive marker "to".

Predictive model
The features we used are:

1. Presence of a progressive verb followed by the infinitive
2. Presence of infinitive marker "to" before the infinitive

Analysis
The naive Bayes multinomial classifier performed the best with an F1-measure of 0.967. We encountered exceptions with words like "saying". An example of such a sentence would be "He was saying play". Most of our false positives were due to sentences such as this. We considered a subset of utterances in which the infinitive was used along with the progressive verb. The missing infinitive marker "to" is also found in other utterances such as "I would love to swim" in which case we have two verbs that are in the base form - "love" and "swim".

4.7 Combining the Classifiers

Finally, we perform sentence level binary classification - does the sentence have a grammatical error? Since an utterance can contain more than one error, we serially apply the binary classifiers that we described above for each error category. If any one of the classifiers reports an error in the utterance, we flag the utterance as having a grammatical error. For evaluation, as long as the utterance had any grammatical error, we considered the decision to be correct. As we can see from Table 3, the best result for detecting the overall errors was obtained by serially applying the classifiers that used the features that were not rule based.

5 Conclusions and Future Work

In this paper, we described a study of grammatical errors in child language transcripts. Our study showed that a higher percentage of children with LI made at least one mistake than TD children on most error categories. We created different systems including rule based systems that used parse tree template matching and classifiers to detect errors related to missing verbs, subject-auxiliary agreement, subject-verb agreement, missing infinitive marker "to", missing copulae and wrong usage of -ing participle. In all cases, we had a recall higher than 84%. When combining the classifiers to detect sentences with grammatical errors, the classifiers that used features other than rules performed the best with an F1-measure of 0.967.

The error categories that we detect at present are restricted in their scope to specific kind of errors. In future, we plan to enhance our systems to detect other grammatical errors such as missing articles, missing prepositions and missing main verbs in utterances that do not have an auxiliary verb. Furthermore, we will investigate methods to address issues in child language transcripts due to incomplete utterances and disfluencies.

At present, we treat sentences that conform to formal English language as correct. We could enhance our systems to look at dialect specific constructs and grammatical errors made across different demographics. For example, African American children have a different dialect and do not always follow the formal English language while speaking. Therefore, in the context of detecting language impairment, it would be interesting to see whether both TD children and LI children make the same errors that are otherwise considered the norm in the dialect they speak.

Acknowledgments

The authors thank Chris Dollaghan for sharing the Paradise data, and Thamar Solorio for discussions. This research is partly supported by an NSF award IIS-1017190.

References

Mari I. Bowden and Richard K. Fox. 2002. A Diagnostic Approach to the Detection of Syntactic Errors in English for Non-Native Speakers. Technical report, The University of Texas-Pan American.

Andrew Caines and Paula Buttery. 2010. You talking to me?: A predictive model for zero auxiliary constructions. In *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground*, pages 43–51.

Eugene Charniak. 2000. A maximum-entropy-inspired parser. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*, pages 132–139.

Lloyd M. Dunn. 1965. *Peabody picture vocabulary test*. American Guidance Service Circle Pines, MN.

Jens Eeg-Olofsson and Ola Knuttson. 2003. Automatic grammar checking for second language learners-the use of prepositions. In *Proceedings of NoDaLiDa*.

Keyur Gabani. 2009. Automatic identification of language impairment in monolingual English-speaking children. Master's thesis, The University Of Texas At Dallas.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18.

Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault. 2010. Automated Grammatical Error Detection for Language Learners. *Synthesis Lectures on Human Language Technologies*, 3(1):1–134.

John Lee and Stephanie Seneff. 2006. Automatic grammar correction for second-language learners. In *Proceedings of INTERSPEECH-2006*, pages 1978–1981.

John Lee and Stephanie Seneff. 2008. Correcting misuse of verb forms. In *Proceedings of ACL-08:HLT*, pages 174–182.

Jack L. Paradise, Thomas F. Campbell, Christine A. Dollaghan, Heidi M. Feldman, Bernard S. Bernard, D. Kathleen Colborn, Howard E. Rockette, Janine E. Janosky, Dayna L. Pitcairn, Marcia Kurs-Lasky, et al. 2005. Developmental outcomes after early or delayed insertion of tympanostomy tubes. *New England Journal of Medicine*, 353(6):576–586.

Mabel L. Rice, Kenneth Wexler, and Patricia L. Cleave. 1995. Specific language impairment as a period of extended optional infinitive. *Journal of Speech and Hearing Research*, 38(4):850.

Kenji Sagae, Eric Davis, Alon Lavie, Brian MacWhinney, and Shuly Wintner. 2007. High-accuracy annotation and parsing of CHILDES transcripts. In *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition*, pages 25–32.

Hollis S. Scarborough. 1990. Index of productive syntax. *Applied Psycholinguistics*, 11(01):1–22.