PARSING WITH A SMALL DICTIONARY FOR APPLICATIONS SUCH AS TEXT TO SPEECH

Douglas D. O'Shaughnessy

INRS-Telecommunications
3 Place du Commerce
Nuns' Island, Quebec H3E 1H6 Canada

While the general problem of parsing all English text is as yet unsolved, there are practical applications for text processors of limited parsing capability. In automatic synthesis of speech from text, for example, speech quality is highly dependent on realistic prosodic patterns. Current synthesizers have difficulty obtaining sufficient linguistic information from an input text to specify prosody properly. When people speak, they often use the syntactic structure of the text message to determine when to pause and which words to stress. Previous work on natural language processing generally assumes access to a large dictionary so that parts of speech are known for virtually all possible words in an input text. However, some practical natural language systems are constrained to limit computer memory and access time by minimizing dictionary size. Furthermore, in most published text to speech work, the parsing problem is only briefly mentioned, or parsing occurs on a local basis, ignoring important syntactic structures that encompass the entire sentence. The system described here recognizes function words and some content words, and uses syntactic constraints to estimate which words are likely to form phrases. This paper is the first to report on parsing details specifically for speech synthesis, while using only a small dictionary (of about 300 words).

1 INTRODUCTION

Parsing a sentence requires information about the parts of speech of its words. Previous work on natural language parsing has generally assumed that parts of speech are known for all words in an input text (Marcus 1980, Grishman 1986). For example, the EPISTLE system (Jensen 1983, Heidorn 1982) employs a 130,000-word dictionary. Although a small dictionary of 200-300 words suffices for the function words (e.g., prepositions, pronouns), being able to identify nouns and verbs has required much larger dictionaries. Locating the verbs in a sentence is particularly useful to specifying prosody, because pauses often occur immediately before or after a verb group. The system described in this paper recognizes all function words and some content words, and uses syntactic constraints to estimate which words are likely to form phrases. It is compared to similar systems using dictionaries in excess of 2,000 words, which have been only partially described in the literature (Dewar 1969, Bachenko 1986). To the author's knowledge, these latter systems are the only other ones that have attempted parsing on arbitrary text with dictionaries of fewer than 10,000 words. Because the parser described here has access only to a very small dictionary, it cannot exploit many of the advances in parsing in recent years. What is explained below, however, is that accurate parsing need not require large dictionaries.

1.1 SYNTHESIS APPLICATIONS

The input for automatic speech synthesis systems can take several forms. In question-answer applications, a user may access a data base with information stored in non-textual form, e.g., tables or numbers. Such a system can use a very limited grammar in formulating the syntactic structure of the output speech ("concept to speech": Young 1979). In some future systems, the queries may be in the form of speech, and automatic speech recognizers will extract prosody and syntax patterns, which can in turn be of assistance in synthesizing responses.

A more immediate synthesis application is automatic text to speech synthesis (Klatt 1987). The conversion of arbitrary English text to speech is useful in aids for the blind and in general voice response systems. Visually handicapped people (few of whom know Braille) can have direct access to the vast wealth of printed information via an optical character reader and a text to speech synthesizer. Concerning voice response, much information in data bases is in the form of text; with an automatic text to speech system, people could telephone a remote data base and hear a vocal version of the information. The queries must be entered through

 0362-613X/89/010097-108-$03.00

the telephone keypad or via speech of isolated words (where prosody and syntax plays no role), but the output speech can be in the form of sentences.

In synthesis from a text of English sentences, the naturalness and intelligibility of the output speech is highly dependent upon realistic prosodic patterns (O'Shaughnessy 1983a). Current synthesizers have difficulty obtaining sufficient linguistic information from an input text to specify prosody properly. The syntactic structure of the text, in particular, is a major factor in determining where a speaker should pause, which words to stress, and how to use pitch rises and falls. However, the problem of parsing natural English, even using a large dictionary indicating parts of speech for all possible words, is as yet unsolved. English allows many syntactic constructions, which one recognizes when reading a text aloud. Text to speech systems, especially when pronouncing sentences with few punctuation marks, perform much more poorly than humans do. In some systems, the problem is further complicated because the number of entries in the dictionary must be minimized for economy. Such systems usually employ letter to phoneme rules, and a small dictionary to pronounce words for which the rules are inadequate. For certain words, knowledge of their syntactic role is imperative for proper pronunciation; e.g., *refuse*, *wind*, *lives*, *separate* use different sounds depending on whether they act as noun, verb, or adjective.

Very little work on parsing sentences for speech synthesis purposes has been reported. This paper is the first to give parsing details specifically for synthesis while using a dictionary of fewer than 300 words. In most other references, the parsing problem is only mentioned in passing (Flanagan 1970; Coker 1973; Klatt 1987). The most documented system, MITalk-79 (Allen 1987), uses a large dictionary and treats parsing only on a local basis, ignoring important syntactic structures that encompass the entire sentence.

Restricting the dictionary to a few hundred entries limits the ability of a parser to correctly analyze all texts. For text to speech, however, it is unnecessary to have a complete parse of the text to be spoken. The dictionary and pronunciation rules must be powerful enough to avoid mistakes in the translation of letters into phonemes, of course. But syntactic structure is useful mostly in specifying prosody, e.g., when to pause, which words to stress, and whether to raise or lower pitch at the end of a sentence. Syntactic information sufficient to specify prosody rarely requires a complete parse. Positions of major syntactic boundaries and identification of stressed words are of major concern. Confusions between nouns and adjectives, for instance, have little bearing on prosody. Using a flexible parser, moreover, minimizes the chance of meeting an unparsable text (Weischedel 1980). A parsing failure in synthesis systems is only serious if it results in an incorrect prosodic assignment that adversely affects the intelligibility or correct interpretation of the output speech. While a local parsing error in one part of a sentence may lead to errors elsewhere in the sentence, many minor errors that occur in our parser due to use of a small dictionary have little effect on the important aspects of the global sentence parse.

1.2 SYNTAX AND PROSODY

The relationship between a text and its prosody is complex. Speakers vary pitch, duration, and intensity (the aspects of prosody) primarily to highlight certain words for the listener and to partition the utterance into short segments for easier perceptual processing (O'Shaughnessy 1983b). Speakers tend to pause at major syntactic boundaries, but the frequency and duration of the pauses also reflect the length of the phrases (measured by the number of words or syllables) between pauses (Gee 1983). Syntactic boundaries are also often cued, in addition to pauses, by a pitch rise and lengthening of the final syllable prior to the boundary. Speakers have much freedom in choosing where and how long to pause and which words to emphasize; such choices are motivated by their desire to communicate meaning to a listener. Thus the semantics of a text is as important for specifying prosody as its syntactic structure (Selkirk 1984). Unfortunately, automatic semantic analysis of arbitrary text is very difficult and not feasible for text to speech (for concept to speech synthesis, on the other hand, semantics may be more readily obtained). Since syntactic structure correlates well with prosody in speech spoken at a normal rate, without emotional and other contextual influences, the parse of a text is a feasible alternative to semantic analysis for text to speech prosody.

Besides indicating likely pause locations, the other major way that syntax influences prosody is that speakers stress important words, i.e., words that are unexpected by listeners and add most to the information content of an utterance. Thus most function words are not stressed. A dictionary that identifies the function words can cue a synthesizer to stress all other words. The amount of stress a word receives is proportional to its importance (or its unexpectedness). Small function words occurring in syntactically restricted (and thus somewhat redundant) positions rarely have semantic importance. As far as part of speech is concerned, however, the words with the greatest stress are included in our dictionary: sentential adverbs, *not*, modal verbs, quantifiers, and interrogative words tend to be more stressed than nouns, verbs, and adjectives (O'Shaughnessy 1983b). Among the unidentified words, there is no large variance in stress due to part of speech, and therefore no need to further specify them for stress purposes. Due to the large prosodic effects at pauses, however, we must try to specify the syntactic role of such words sufficiently to find pause locations.

Syntactic structure also affects prosody in other ways besides pausing and stress (O'Shaughnessy 1979). Pitch rises sharply at the end of a question asking for a

yes or no response. Parenthetic expressions, often offset from the main part of a sentence by commas or parentheses, are uttered with reduced pitch. Vocatives can be distinguished from appositives by different pitch patterns. Parsing information is useful to a synthesizer to handle all of the above effects.

2 Segmenting Sentences into Phrases

For text to speech, the primary task for a parser is to segment a sentence into phrasal units each containing a few words. Such units often act as prosodic (intonational) groups. Pauses are usually restricted to come at the boundaries of these units, and the final word (among others) in each unit is usually stressed. Determining the higher-level syntactic structure that links these phrasal units together is often more difficult.

Thus parsing for speech synthesis can employ two strategies: local and global. The local strategy typically operates first and goes left to right through each sentence, i.e., as the words enter the system. For real time applications, it may be important to output parsing results even before the final words of a sentence are available. Since a global strategy may require examining as much as the entire sentence, it may revise the parse of the early parts of a sentence as later words are analyzed. The global analysis should also attempt left to right (real time) analysis; this is adequate for most sentences, but ones with complex syntactic structure (e.g., unpunctuated subordinate clauses) often require examination of the entire sentence for a correct parse. Sentence-final punctuation (! ?) can significantly affect the sentence's prosody (and parse, to a lesser extent). In the case of long sentences with such final punctuation, however, the prosodic changes due to the punctuation primarily affect only the last clause.

Locally, our parser groups words likely to act as prosodic units. This means composing various phrases (perhaps smaller than traditional linguistic units) out of component words: 1. noun group (NG), which consists of a noun and its immediately preceding words (e.g., article, quantifier); 2. verb group (VG), which consists of a verbal word optionally preceded by modal and auxiliary verbs; 3. prepositional phrase (PP), which consists of a preposition followed by a NG; 4. adjectival phrase (AdjP), which consists of an adjective, possibly preceded by an adverb; 5. adverbial phrase (AdvP) (see Appendix 1). (This list is similar to that in Bachenko 1986.) For the purposes of local parsing, NGs and VGs are more useful units than noun phrases (NPs), which consist of an NG followed by PPs or AdjPs, and verb phrases (VPs), which consist of a VG followed by its complement(s). Pauses are rare within word sequences corresponding to the five basic phrasal units noted here, but often occur within an NP or VP. To help locate phrase boundaries, the parser exploits constraints on word order in NGs and VGs; when normal order appears to have been violated, it is likely that a phrase boundary has occurred at the point of deviation.

The problem of sentence segmentation is assisted by punctuation marks (e.g., commas), which often occur at clause boundaries. However, many sentences have little internal punctuation. The word sequences that colons and semicolons delimit can be treated as sentences for prosodic purposes. Left marks (quotes, parentheses, brackets, braces) act as phrase-introducing marks, and corresponding right marks terminate phrases. Both dashes and commas tend to partition the sentence into clauses and phrases, and are likely places for pauses and prosodic marking, especially as the lengths of the phrases they delimit increase in size. However, commas are not restricted to delimiting major syntactic units. In lists of two or more units (of similar syntactic identity), commas are often internal to major phrases (e.g., "foxes, mice, and birds" forms a single NP). Although the words just prior to such commas are often prosodically marked with pitch rises, pauses are usually reserved for boundaries between long or major phrases. Furthermore, lists of words containing commas do not always employ a coordinate conjunction (e.g., "a slimy, round, large, red fish"). It often makes little difference to the prosody of such phrases whether the commas are present or not; thus one cannot treat each comma as a syntactic boundary.

3 Preprocessing (Text Formating)

In text to speech, special processing is needed for text entries not in word form (e.g., digits, abbreviations), which must be converted into corresponding words. This preprocessing can assist in grouping words. Abbreviations often represent words that are closely linked to adjacent words; e.g., measurement abbreviations (sec., mi., oz.) are usually preceded by a numeral or quantifier, forming a NG. Four classes of abbreviations depend on the direction of linkage with adjacent words (examples are given in parentheses): 1. left (Jr., in., Blvd.); 2. right (Mr., Mrs., Prof., Fig.); 3. either (Tues., Dept., St.); 4. both (vs., cu.). Virtually all abbreviations form NGs with immediately adjacent words (e.g., Main St., Mr. Jones), although some in the fourth class may link words on a broader scale (e.g., "vs." can link two arbitrarily long NPs).

Other text preprocessing of use to parsing concerns hyphens, capital letters, and contractions. Hyphenated words (e.g., tongue-in-cheek) are treated as nouns, unless all their components are numerals (e.g., forty-one), which fall into the numeral category. A string of capitalized words is considered to be a phrasal unit, because it is likely to be a NG and be spoken without a pause. Words consisting entirely of capital letters are usually acronyms and, like digit sequences, act as nouns (or adjectives). Most contractions are uniquely converted to words ('ve → have); for others the conversion is not unique but the ambiguity has no effect on parsing

('d → had or would). Most contractions involve auxiliary verbs and have minimal prosodic effects. On the other hand, *-n't* (= not) is important prosodically since the preceding syllable becomes stressed. The contraction *'s* can either be a verb (is, has) or act as a possessive adjective (John's); thus the parser must allow two possibilities. Heuristic rules help here: 1. after a pronoun, *-'s* is verbal; 2. a possessive contraction is usually followed by an adjective or a noun; 3. the verbal *-'s* usually precedes a verb participle (e.g., an -ing or -ed word). Confusions here do not have severe prosodic effects because pauses do not occur right after *-'s* contractions.

4 Word Dictionary and Procedures for Each Part of Speech

The parsing dictionary consists of about 300 words, each labeled with 1–3 possible parts of speech. About 50 of the words have 2–3 possible classifications (e.g., "it" can be either a subject or object pronoun; "more" can act as noun, quantifier, or adverb). For words with multiple syntax possibilities, the most probable is tried first, and the others are only used in case of parsing failure.

The parts of speech that the parser employs can be grouped into classes, which are subdivided according to the useful parsing features that distinguish words. The dictionary contains about 60% function words and 40% content words. The largest classes of function words are the prepositions and conjunctions, each having about 13% of the dictionary words. They are followed (in order of decreasing size) by auxiliary verbs, pronouns, numerals, quantifiers, and articles. The dictionary content words are dominated by the adverbs (25% of the dictionary), with common verbs making up most of the rest. The remaining thousands of words that are not in the dictionary fall into four classes: noun, adjective, verb, and adverb. Adverbs not in the dictionary end in *-ly* and are thus identified by their suffix. Hence words not recognized by the dictionary are limited to act as verbs or nouns. (In terms of parsing for prosody, little is lost by treating adjectives as nouns, since they often occur interchangeably in similar positions, e.g., as a complement after a verb or as modifiers in an NG.)

The role of the dictionary then is to specify the syntactic functions of all words that belong to small classes of words. A word dictionary of 300 entries, augmented by a suffix analyzer (described in Section 5) using fewer than 60 suffixes, is sufficient to identify all words except those acting as nouns, adjectives, and verbs. These latter classes are open and contain a theoretically unlimited number of words. The power of the parser can be increased by including some of these words (e.g., verbs with irregular endings or which take two objects), but the dictionary rapidly increases in size in such cases with only limited benefits for prosody.

At the local parsing level, the system accepts each new word (from left to right), searches the dictionary for a match, and, if successful, attempts to link the new word to the immediately preceding words to form syntactic phrasal units (NGs, PPs, VGs, AdvPs, AdjPs). When the new word is syntactically incompatible with prior words (for reasons described below), a new phrasal unit is started. The procedure is detailed below and is organized according to the part of speech of each word found in the dictionary. (In the following discussion, word examples are given in parentheses.)

1. A **preposition** introduces a PP and thus starts a new phrasal unit (which may be grouped later, at the global level, with a prior NG and ensuing PPs to form a NP). Some prepositions (despite, besides) may precede gerunds, while others (about) can precede *to* + infinitive (where the sequence then forms an infinitival phrase), and those in a third subclass (instead, because) can merge with an ensuing "of". By distinguishing these subclasses, the parser can better decide whether or not to link a preposition with ensuing words; e.g., when a non-gerund-preceding preposition such as "under" is followed by a gerund, a syntactic boundary separates the two words (which indicates that "under" either ends a clause or is acting as an adverb). One preposition is special: "to" can be followed by either an NG or an infinitive (an infinitive is assumed if a content word follows "to").
2. A **conjunction** also indicates the start of a new phrasal unit. For those that introduce dependent clauses or phrases (when), no link is made with preceding words; for coordinate conjunctions (and, but) the global parser later attempts to merge phrasal units adjacent to the conjunction into a larger unit. Some conjunctions (unless) may precede gerunds and participles, while others (because) can only precede clauses.

While relative pronouns (where, that) and adjectives (whose) can act as conjunctions in starting clauses, such clauses function as NPs (and serve as a subject or object in a clause for the higher-level parse), whereas clauses introduced by other conjunctions are not directly linked to adjacent words. The presence of a Wh-word at the start of a sentence (perhaps right after a preposition) indicates that the sentence is a Wh-question if a verb immediately ensues (e.g., With whom does he eat? vs. What he eats is fish.). When meeting the word "as", the parser looks for an ensuing "as" to link into a larger phrase (e.g., as blue a fish as I could find).

3. A **pronoun** usually acts as a one-word NG and thus, at the local level, can only link with an immediately prior preposition (exception: "we", "you" can act as a quantifier—e.g., you blue meanies). Certain pronouns only function as subject NGs (I, we) and indicate that the

ensuing words form a VG. Others behave as object NGs (us) and indicate that the prior word is either a verb or a preposition. Reflexive pronouns (words ending in -self) behave prosodically like adverbs and are stressed; it is unnecessary to see if a reflexive pronoun matches the preceding word, since reflexives tend to act prosodically as sentential adverbs, getting their own stress contour (e.g., in "Sue hit John himself/herself", whether the pronoun links to "Sue" or "John" makes little prosodic difference). Some pronouns act as adjectives, which either must be followed by the rest of a NG (possessive pronouns—our), stand by themselves (ours), or have both options (demonstrative pronouns—those).

4. A **quantifier** generally starts an NG, sometimes jointly with an article (such a, a little, the other). Certain quantifiers indicate number (singular—much, every; plural—many, some), which is useful in locating the end of the NG (e.g., in "Every cat walks home . . .", the word "walks", potentially a plural noun, cannot be part of the subject). Except for predeterminers (almost), only one quantifier can occur in a NG; thus a sequence of two quantifiers is broken into two NGs. After a comparative quantifier (more), the parser looks for "than" to form a larger unit (e.g., *more* NG *than* [NP, S] acts as a NP).
5. An **article** always starts a new NG (except after certain quantifiers). In the case of *a/an*, the NG is singular and thus should not terminate with a plural noun (exceptions: a great many, a [hundred, thousand, million]).
6. **Numerals** form a class of words of marginal utility to the parser. The cardinals (one, ten) are useful for specifying the number of its NG (e.g., in "After one night dogs swam home . .", "dogs" cannot be part of the initial PP), but the ordinals (first, third) only serve to possibly indicate NG boundaries (e.g., In summer third-rate movies are . . .). A cardinal numeral is the last function word of a NG (exception—one another); e.g., in "With those two some men win", the quantifier "some" starts a new NG.
7. **Adverbs** form a class of content words, but, excluding words ending in *-ly*, it is the smallest such class and is feasible to be included in the dictionary. Identifying adverbs is especially useful for a prosodic parser because they tend to be strongly stressed and their syntactic functions help parse the sentence. Each adverb has one of three roles: 1. as a sentential adverb (seldom), which can appear virtually anywhere in the sentence, and thus should be ignored when looking for syntactic structure; 2. modifying (and following) a verb (aloud), which helps locate one-word VGs; and 3. modifying an adverb or adjective (quite), which labels the ensuing word. For example, in "has actually eaten", the adverb "actually" should be ignored as part of the VG; in "Large fish swim away", "swim" is identified as a verb because it precedes the class-two adverb "away"; in "Very hungry people eat food", "hungry" must be an adjective (following the class-three adverb "very") and thus "people" cannot be a verb. The adverb "not" has other parse functions: except when following an auxiliary verb, "not" starts a phrasal unit, either an NG (e.g., At the beach, not swimming is dumb) or a verb complement (e.g., Are blue fish not cold?).
8. **Auxiliary verbs** (forms of be, have, do) and **Modal verbs** are very useful in parsing because they initiate a VG (and thus terminate a preceding phrase) and often indicate the number of the subject; e.g., in "The fear animals show is temporary", "animals show" can be identified as a subordinate clause because "is" must have a singular subject and the plural "animals" must be the subject of a relative clause.
9. A few common **Verbs** (made, read) are included in the dictionary because identifying each clause's verb is important for prosody. Most useful are past tense verbs (kept) that do not end in *-ed*, because these irregular verbs appear frequently and are not easily identified by suffix analysis. The 2,000-word system noted earlier (Bratley 1968) deviates significantly from our system by including a large number of verbs, with each entry noting how many objects (0, 1, or 2) are expected to follow the verb. There are several reasons we do not use a list of intransitive verbs (die) and two-object verbs (gave, offer): 1. their large number; 2. an indirect object is optional—thus the parser cannot rely on its presence; 3. it is often difficult to tell if a sequence of unidentified words after a VG forms one or two NGs. Thus our parser allows for 0–2 NGs after each verb.
10. A few common **Nouns** are in the dictionary to aid in specifying number. Virtually all plural nouns end in *-s*, but a few common nouns do not (people, men, women). Thus our parser assumes that unidentified words not ending in *-s* are singular.

5 Morph Decomposition and Dictionary

A 300-word dictionary suffices to recognize almost half the words in general text, but the syntactic role of most content words remains unspecified. Since many English content words end in suffixes that help identify their part of speech, it is useful to try to classify words not found in the dictionary by their endings. The MITalk

system (Allen 1987) decomposes each such word into all its component "morphs" (prefixes, roots, and suffixes). Our simpler approach just looks at the final letters in these words to locate likely suffixes (-ness, -able). A list of about 60 suffixes (ordered from longest to shortest) is compared to the endings of such words. Each suffix is associated with its most likely part of speech (as determined from an analysis of English words). Only words with two or more syllables are examined for possible suffixes (except for the *-s* suffix), since virtually all words with suffixes have a root morph containing one or more syllables, to which the suffixes attach. If a letter to phoneme translation is unavailable, a simplistic syllable counter can suffice in which a series of adjacent letters from the set [a,e,i,o,u,y] (except word-final -e) counts as one syllable. (While this miscounts some words, it suffices for finding suffix eligible words; e.g., one-syllable words like "worked" are accepted, and words like "fable" and "size" are rejected.)

If a word ends in *-s*, special rules are invoked (unless the preceding letter is i, u, or s): the *-s* is stripped off, and the suffixes are re-examined on the shortened word. For this second pass, any suffix ending in *-y* is modified to end in *-ie* (e.g., to label "identifies" as a verb, the verbal suffix *-ify* must be changed to *-ifie* in the presence of the final *-s*). If a suffix is found for the shortened word, the corresponding part of speech applies, but the word is noted as being either plural (in the case of a noun) or third-person singular present active (for verbs). If no further suffix match is found (other than the original *-s*), then the two possibilities (plural noun and singular verb) are retained. While the noun/verb ambiguity remains (before examining context), words ending in *-s* are very useful when examining adjacent words for number agreement. Two passes are made through the suffix dictionary only if the word ends in -s. This is different from MITalk word decomposition, which, using a large morph dictionary, continues to strip off as many affixes as possible, until the root form is left. To the extent part of speech can be determined from decomposition without a large dictionary, analysis of the last suffix (two suffixes when the word ends in -s) is sufficient.

Many word terminations uniquely specify a part of speech. Words whose final letters match one of the following suffixes are considered identified for syntax purposes: nouns (-ity, -or, -ship, -time, -ness, -sm), adjectives (-ous, -ful, -less, -ic), verbs (-sist, -the, -ify), gerunds (-ing), and numerals (-teen). Other word endings are probable indicators of part of speech (e.g., -ment →noun). Several tentative verb endings, primarily past tense forms (-ed, -ught, -ung), are included because past tense verbs do not provide parser assistance through number agreement rules (past tense verbs can accept both plural and singular subjects).

6 Parsing Algorithm

The system uses a bottom-up parser based on a context free grammar, with constraints on permissible syntactic groupings (see Appendix A). As in an augmented transition network (Woods 1970), the algorithm of IF-THEN procedures involves transitions between states and their consequences. The consequences of the conditional actions of the rules involve the gradual construction of a parse tree. The states of the network correspond to different syntactic contexts and different stages of parser tree development; e.g., as each word is examined, a transition is taken out of the state specified by the prior words, depending on the part of speech of each new word. Each state has a possible outgoing transition for an unknown part of speech, to handle the fact that many words are classified only as being "content words." For a recognized word with 2–3 possible parts of speech or a word with a tentative suffix, the most likely transition is taken and backtracking occurs if an inconsistency is met or no successful parse results by the end of the sentence (the "depth first" approach).

Other parsers, with access to a large dictionary specifying part of speech and other attributes for virtually all words, can operate with a tight, completely detailed grammar and produce complete parse trees. This may be necessary for natural language understanding, but is not needed for specifying prosody. Our restricted dictionary (especially the lack of any knowledge of attributes for virtually all nouns and verbs) forces us to weaken the grammar of English and to output incomplete parse trees. For prosody, it suffices to label phrases and locate their boundaries.

6.1 Control Procedure

For each new word in the input text, a procedure (as described in Section 4) corresponding to the (possibly tentative) part of speech is called to: 1. combine the word with immediately prior words to form a local phrasal unit (perhaps renaming the unit as the new word is added), or 2. decide that the word starts a new phrasal unit. If the word is not identified by the dictionaries, a tentative part of speech is estimated from context, and the same procedure is followed. As each new phrasal unit is started (case 2), the parser operates at the global level to link the previous phrasal unit to earlier units, to identify clausal units (e.g., main and dependent clauses) and their components (e.g., subject and object NPs). To group words together, we exploit restrictions on word order in phrases, as well as on word and number agreement in adjacent phrases.

If the system finds an inconsistency (e.g., no legal parse according to the grammar, or a violation of the syntactic constraints), it rejects the current parse and backtracks to the last tentative decision concerning either grouping of words in phrasal units or choice of part of speech. An alternative is chosen and the parse

continues from that point. Given the large degree of ambiguity caused by the small dictionary and words with several syntactic roles, it is more efficient to proceed depth first using the most likely choices, than to process possible paths in parallel. This is especially true given that we desire only a parse tree sufficiently detailed to predict prosody; many of the trees produced by parallel paths would correspond to equivalent prosodic patterns. The depth first approach is least efficient when a tentative decision early in a long sentence must be reversed; e.g., in "That fear men have is stupid", deciding that "that" is a demonstrative only comes after analyzing the rest of the sentence.

6.2 CASE OF AN UNIDENTIFIED WORD

To identify the part of speech of most nouns, adjectives, and verbs, context analysis must be invoked. One aspect of exploiting context is ensuring that words are consistent in number within a NG and between a subject NG and its VG; e.g., if a subject NG is plural, an ensuing one-word verb should not end in *-s*. Number rules are useful since the parser is often faced with a sequence of unidentified words: 1. If none of the words ends in *-s*, they are likely to form a single NG; 2. if the last one ends in *-s*, it could be a verb, with the other preceding words forming a singular subject NG, or the entire sequence could be a plural NG; 3. if the penultimate word ends in *-s*, the last word is likely to be a verb, with preceding words forming a plural NG.

A major problem is locating phrase boundaries that are not marked by function words. NGs that commence with a function word are easily found, but some NGs consist of only adjectives and nouns. An especially difficult, yet prosodically important, case concerns locating the boundaries of embedded clauses not offset by commas. For example, in "The turtles (that) men see swim well", the parser could see a dictionary output of Article + Unknown + Unknown + Unknown + Unknown + Adverb (assuming that "that" is not present and "men" is not in the dictionary). In this case, among the unknown words only the first ends in *-s*, so agreement rules try to label "men" as a verb, leaving "see swim" as a NG. By including the (relatively few) common plural nouns that do not end in *-s* (e.g., men, women) in the word dictionary, however, the situation here can be clarified. If "men" is found as a plural noun, then the ensuing word (see) is marked as a verb, and the following word (swim) is labeled as the verb for "the turtles" with "men see" as an embedded clause.

Number agreement rules are invoked primarily when plural words are identified. Inside subordinate clauses and phrases, a singular NG may often precede an active verb not ending in *-s*; e.g., I insist that he eat.

Specifying the part of speech for an unidentified word X is based primarily on the role of its immediately preceding word W. If X starts a sentence, it is called a noun unless the immediately following word is an introductory word; this latter case is that of an imperative, where X is a tenseless verb. X is also called a noun when: 1. W is an article, quantifier, demonstrative, numeral, adjective, preposition, gerund, subordinate conjunction, or "whose"; 2. the sentence starts with *that* X; or 3. X is preceded by a two-word infinitive phrase (i.e., *to* + a singular unknown word). X is called a verb after "*who*" or a sentential adverb. The remaining cases depend upon X's number. A singular X (i.e., not ending in -s or ending in -is, -ss, -us) is called a verb after: 1. an auxiliary or modal verb (He will work) (unless the sentence starts with such a verb—Can fear rule?); 2. a sequence of verb + coordinate conjunction (He ate and ran); 3. a subject pronoun (*He ran*), 4. a plural NG (The bells ring); otherwise, it is named a noun (blue cheese). A plural X is called a verb after: 1. a singular NG (a rat smells); 2. a relative pronoun (what eats); 3. a singular subject pronoun (He swims); otherwise it is named a noun (communications engineers). In several of these cases, the choices are successful only for a majority of sentences, because a universally correct choice is impossible before examining later words (if then). Thus, every time a word unidentified by the dictionary is tentatively labeled here, the alternative choice (noun or verb) is stored in a stack for use if a later parse failure causes backtracking to this word. For example, in "What cats do is unclear", "cats" is first labeled a verb, but then the ensuing identified verbs "do is" force "cats" to be renamed a noun because they each need a subject.

6.3 INDEPENDENT PHRASES

Sentences often have phrases that do not directly modify the subject, verb, or object of a clause. Instead, they are either parenthetical (e.g., he said, you know) or modify a clause as a whole (AdvPs). The parser easily handles such expressions when they are offset by punctuation. Without punctuation, however, the problem is more difficult but necessary to solve, since such expressions usually have their own prosodic grouping.

A common independent phrase is a temporal adverbial—a NG introduced by a function word, where the final noun deals with time (e.g., last week, three times a month, next Tuesday). The expression tends to act as a sentential adverbial, occurring at any of a number of locations in a sentence. One way to identify these expressions is to list in the dictionary all nouns dealing with time (e.g., second, day), but the list is apparently large (e.g., semester, period, etc.). We avoid that approach, since these expressions are readily isolated (although not so easily labeled) in virtually all cases since they commence with a function word that causes it to be recognized as a NG. Such a NG can be confused with a subject or object NG, however. The only risk is that, in a sentence with short phrases, a temporal adverbial might be syntactically merged with an adjacent VG as its subject or object (and as a result, not be prosodically separate). Since short temporal adverbials are not always isolated prosodically, such a risk is not

serious. The only case where a temporal adverbial does not commence with a function word, and thus risks being merged with a preceding NG, occurs when the time noun is plural and is followed by a comparative adjective (e.g., days later); thus the parser looks for such situations in possible long NGs.

6.4 INTRODUCTORY FUNCTION WORDS

The word "that" can act as a noun, a demonstrative adjective, or (most often) a relative pronoun. The parser tries the most common role first, where "that" must be followed at least by a subject NG and a VG, forming a subordinate clause. Unlike *that*-clauses, clauses introduced by other relative pronouns (Wh-words) may follow a preposition or may act as the subject or object NG of the clause. Such a clause may end with a preposition if a subject NG precedes the VG (e.g., in "Who[mever] John gave it to was not clear").

The other use of clauses with "that" or a Wh-word is following (and modifying) a NG. In such cases, "that" functions exactly like the Wh-words, except that "that" is often deleted (The man (that) John gave it to was here). Such clauses must directly follow the NG they modify; adverbials and other parenthetic expressions are not allowed. Finding the end of such clauses (and the beginning, in the case of a deleted "*that*") is an important task for the parser. In the example above, the sequence "to was" forces a syntactic boundary, and the parser searches for a moved constituent, which strongly suggests the presence of a subordinate clause; the unlabeled words "John gave" would then be interpreted as a subject NG followed by a verb.

Certain NG introducers impose a number constraint on the NG. A NG starting with "a(n)", "this", or "every" cannot end with a plural word (exception: if "few" or "great many" follows "a"). Thus, if the parser encounters a series of unidentified words after a singular NG introducer, it assigns the first plural word to an ensuing phrasal unit (e.g., in "When I bought this four people gasped", a boundary is placed between "this" and "four"). (If a plural word ensues directly, it is treated as an adjective, e.g., "a communications engineer".) When a plural numeral occurs in a singular NG (e.g., a blue four-foot ladder), then the parser links the numeral to the next word to act as an adjective.

6.5 GLOBAL GROUPING

Each sentence is assumed to have a subject (perhaps implied, in the case of imperatives) and a verb. Normal order is subject NG, VG, and 0–2 object NPs. (PPs and AdvPs are free to occur at any NG or VG boundary.) Any of these elements can consist of a set of coordinated units (e.g., the VG could be "ate, drank, and slept"). The parser marks any deviation from normal order as a potential syntactic boundary; e.g., the initiation of a second NG during the course of an apparent NG is a cue to a possible clause boundary. In general, deviations from the expected order of events (e.g., a PP initiating a clause, rather than after a NG or VG) are occasions when a speaker tends to mark syntax prosodically.

Short sentences may not greatly mark NP–VG–NP boundaries prosodically, but long sentences usually do, depending on the relative sizes of the phrases. In a long NP, pauses are likely to occur at NG—PP boundaries. For coordinated units in a phrase, pauses are more likely at the coordination points (i.e., at punctuation and/or just before conjunctions). Adverbials, especially AdvPs of more than one word, tend to function as separate prosodic units.

An indirect object NG (if present) usually immediately follows the VG (without interruption by an AdvP). If an object NG is a simple pronoun, it attaches prosodically to an adjacent phrase; if two object NGs (other than pronouns) follow a verb, each will be assigned its own prosodic group.

6.6 COORDINATED WORDS AND PHRASES

One syntactic structure with strong stress effects is that of parallel contrast, where two (or more) concepts are contrasted in structures either joined by coordinate conjunctions or in a list separated by commas; e.g., in "John ate fish, while Joe ate beef", the repeated verb is less stressed than normally and the other content words, paired in parallel clauses, are more stressed. This parallelism is sometimes extended to the point where the repeated word may be dropped (i.e., gapping may occur). Such ellipsis can drastically alter prosody (e.g., inserting a pause after "Susan" in "John bought fish, and Susan juice"). The most common form deletes the verb from the second of two conjoined clauses. It may be identified by the parser in cases of coordinated units where the first unit is a clause and the second unit lacks a verb.

In cases of conjoined phrases (i.e., ABC . . . GH conjunction JK . . . PQ—where the letters represent words), the parser assumes local coordination and searches for the smallest units to link. After eliminating parenthetic expressions and adverbials set off by punctuation, the parser considers linking phrases up to the nearest punctuation. To link, the phrases must have the same general structure. Thus, if J is a preposition, the parser searches to the left for a preposition to match. If the conjunction links two prepositions (i.e., H is a preposition) (e.g., "He went into and through the house"), the prepositions are stressed, as part of the general process of stressing the parallel constituents in coordinated phrases. If ABC . . . GH and JK . . . PQ both form clauses, the conjunction is viewed as linking the two clauses (the same holds when JK . . . PQ starts with a VG, under the assumption that the implied subject is common to both clauses).

English has certain word sequences which assist in grouping coordinated units. When linking two phrases with "and", the scope of the first unit may be cued by the word "both"; thus the parser, in its search leftward

from "and" looks for a possible "both". Similar actions are taken for the following pairs: *either . . . or, neither . . . nor, whether . . . or not, and not only . . . but.*

6.7 COMPARATIVE STRUCTURES

Phrases involving comparative words (e.g., more, as) are useful to identify because they often group words together prosodically. The discovery of "more" or "less" sets a flag looking for a "than" to link with, setting up a parallel syntactic structure. "More/less" can act either as a noun (by itself), an adverb, or a quantifier-adjective. If "than" ensues immediately, it links tightly to the preceding word; if, on the other hand, some words intervene, a syntactic boundary is indicated right before "than". If "than" appears without a preceding "more" or "less", then the parser searches leftward to link "than" with an *-er* word in the functional role of an adjective (i.e., adjacent to a possible noun, or after a "be" verb). Phrases with "as" are more diverse than "more/less" phrases. "As" followed by a clause is a subordinate clause; "as" followed by a NG is a PP. Either of these word sequences can be preceded by *the same + NP* to form a larger prosodic unit. Another *as*-structure is *such + NP + as* + .. (e.g., such men as these).

6.8 QUESTION ANALYSIS

Most yes-no questions are marked with subject-verb inversion at the start of the final main clause of the sentence. Such a final clause usually starts with a modal or auxiliary verb, which is immediately followed by the subject NG, and then the rest of the VG (if the VG contains more than one verbal). However, virtually any sentence or phrase can be turned into a yes-no question by simply adding a question mark at the end. Thus sentences ending with a question mark are assumed to be yes-no questions, unless an unbound Wh-word (e.g., where, what) is found in the main clause of the sentence. We distinguish bound and unbound Wh-words, since questions in which the Wh-word is bound to a relative clause (e.g., This is where we went?) are yes-no questions. The parser, however, is generally capable of determining whether each Wh-word in the sentence is bound and whether it lies in the final main clause (e.g., in "Did you say who's there?," pitch rises at the end).

Subject-verb inversion is fairly easy to identify. It has the structure of an auxiliary or modal verb, followed by a subject NG, and then by the rest of the VG. The word "not" may occur before or after the subject, and an introductory clause (or AdvP) may precede. In the case of a lack of punctuation (e.g., in "When he came did he eat?"), the unexpected appearance of an auxiliary verb usually helps note the clause boundary. Since a VG can consist simply of an auxiliary, in these yes-no questions the subject NG can be followed immediately by a complement or object (e.g., Is the man blue?). While sentences like "Has John pneumonia?" are theoretically grammatical, they are rare, and question versions of "John has pneumonia" and "John did the work" usually involve the insertion of a conjugated form of "do" (e.g., "Did John do the work?"). As an example of how prosody can be greatly affected by subtle syntax differences: a Wh-word usually cues terminal falling pitch, but in "Did you see what he did?", pitch rises.

7 Comparison to Other Systems

Compared to an earlier 2,000-word system, this parser is almost as successful for the 54 sentences on which it was evaluated (Thorne 1968, Dewar 1969) (see groups 1 and 2 in Appendix B). That parser, not destined for prosodic needs, found two legal parses for four of the sentences that our parser found but one: *The cat adores fish; Fred gave the dog biscuits; he observed the man with the telescope; I dislike playing cards.* A text to speech system requires a single output; e.g., in the first example, we choose the parse where "adores" is the verb (and not "fish"). In a sentence such as "The boy scouts ran", a similar choice would be wrong, but the only way to avoid this is to include many hundreds of verbs in the dictionary. Of the 54 sentences, only five (relatively minor) failures occurred: 1. "Chew gum" was parsed, not as an imperative, but as in "Fear won"; unless "chew" is in the dictionary or the system is biased toward imperatives for short sentences, this ambiguity is not easily resolved; 2. in "He rolled up the bright red carpet", it is impossible to label "up" as adverb or preposition without a significantly complex semantic component; 3. the same comment holds for "Fred gave the dog biscuits", with regard to determining whether there is one or two objects involved. The other two mistakes were ones of incorrectly grouping correctly identified words (see Appendix B).

Our parser was also tested on the 39-sentence (456-word) set of Bachenko (1986) . They claimed only one mistaken parse among the 39 sentences (but did not indicate their parsing output), while our parser made one mistake each in four of their sentences. That our system with only a 300-word dictionary is virtually as successful as ones using dictionaries more than six times larger shows the adequacy of our parser. English has enough syntactic redundancy to allow correct labeling of words' part of speech and location of phrase boundaries using function words and constraint rules as described in this paper. In the relatively infrequent cases where our system makes mistakes, they are rarely of the type that would cause incorrect intonation in terms of misplaced stress, but rather would cause some pauses to occur at secondary syntactic boundaries instead of at major ones.

8 Conclusion

We have described a parser suitable for certain text processing applications where a complete parse may not

be necessary. For example, specifying prosody in a text to speech system basically requires only three things: knowing where to pause (major syntactic boundaries) , which words to stress (distinguishing content and function words), and whether the sentence requires a pitch fall or rise at the end (is it a yes-no question?). The parser uses a 300-word dictionary to identify common words and a set of linguistic constraints to determine likely syntactic structure. The system finds syntactic boundaries where a speaker reading the same text would likely pause, and labels each word with a part of speech with sufficient accuracy to assign proper stress.

References

Allen, Jon; Hunnicutt, M. S.; and Klatt, Dennis 1987 *From Text to Speech: The MITalk System*. Cambridge University Press, Cambridge, England.

Bachenko, Joan; Fitzpatrick, Eileen; and Wright, C.E. 1986 The Contribution of Parsing to Prosodic Phrasing in an Experimental Text to Speech System. *Proceedings of the Association for Computational Linguistics Conference*: 145–155.

Bratley, P. and Dakin, D.J. 1968 A Limited Dictionary for Syntactic Analysis. In: Dale and Michie (eds.), *Machine Intelligence* 2: 173–181.

Carlson, Rolf and Granstrom, Bjorn 1986 Linguistic Processing in the KTH Multi-lingual Text to Speech System. Proceedings of the International Conference of Acoustics, Speech and Signal Processing (ICASSP): 2403–2406.

Coker, Cecil; Umeda, Noriko; and Browman, Catherine 1973 Automatic Synthesis from Ordinary English Text. *IEEE Transactions Audio & Electracoustics* AU—21: 293–298.

Dewar, H.; Bratley, P.; and Thorne, J. P. 1969 A Program for the Syntactic Analysis of English Sentences. *Communications of the ACM* 12(8): 476–479.

Flanagan, James; Coker, Cecil; Rabiner, Lawrence; Schafer, Ronald; and Umeda, Noriko. 1970 Synthetic Voices for Computers. *IEEE Spectrum*. 7(10): 22–45.

Frazier, Lyn 1985 Syntactic Complexity. In: Dowty, Kattunen, and Zwicky (eds.), *Natural Language Parsing*. Cambridge University Press, Cambridge, England 129–189.

Gee, James and Grosjean, François 1983 Performance Structures: A Psycholinguistic and Linguistic Appraisal. *Cognitive Psychology* 15: 411–458.

Grishman, Ralph 1986 *Computational Linguistics: An Introduction*. Cambridge University Press, Cambridge, England.

Heidorn, G.E.; Jensen, K.; Miller, L.A.; Byrd, R.J.; and Chodorow, M.S. 1982 The EPISTLE Text Critiquing System. *IBM Systems Journal 21(3): 305–326*.

Jensen, K.; Heidorn, G.E.; Miller, L.A.; and Ravin, Y. 1983 Parse Fitting and Prose Fixing: Getting a Hold on Ill-formedness. *American Journal of Computational Linguistics* 9(3–4): 147–160.

Klatt, Dennis 1987 Review of Text to Speech Conversion for English. *Journal of the Acoustical Society of America* 82(3): 737–793.

Marcus, Mitchell 1980 *A Theory of Syntactic Recognition for Natural Language*. MIT Press, Cambridge, MA.

Milne, Robert 1986 Resolving Lexical Ambiguity in a Deterministic Parser. *Computational Linguistics* 12(1): 1–12.

O'Shaughnessy, Douglas 1979 Linguistic Features in Fundamental Frequency Patterns. *Journal of Phonetics* 7: 119–145.

O'Shaughnessy, Douglas 1983a Automatic Speech Synthesis. *IEEE Communications Magazine* 21(9): 26–34.

O'Shaughnessy, Douglas and Allen, Jonathan 1983b Linguistic Modality Effects on Fundamental Frequency in Speech. *Journal of the Acoustical Society of America* 74(4): 1155–1171.

Selkirk, Elisabeth 1984 *Phonology and Syntax*. MIT Press, Cambridge, MA.

Tennant, Harry 1981 *Natural Language Processing*. Petrocelli, NY.

Thorne, J. P.; Bratley, P.; and Dewar, H. 1968 The Syntactic Analysis of English by Machine. *Machine Intelligence* 3: 281–309.

Weischedel, Ralph; and Black, John. 1980 Responding Intelligently to Unparsable Inputs. *American Journal of Computational Linguistics* 6(2): 97–109.

Woods, William 1970 Transition Network Grammars for Natural Language Analysis. *Communicatons of ACM* 13: 591–606.

Young, S. J.; and Fallside, F. 1979 Speech Synthesis from Concept: A Method for Speech Output from Information Systems. *Journal of the Acoustical Society of America* 66(3): 685–695.

Appendix A

The bottom-up parser employs the following grammar, which is intentionally loose to allow as much as possible of English text to be accepted. We are not concerned here with detecting ungrammatical sentences, but rather with finding likely syntactic structure.

Abbreviations: adj–adjective; AdjP–adjective phrase; adv–adverb; AdvP–adverb phrase; art–article; aux–auxiliary verb; beaux–"be" verb; card–cardinal numeral; cj–conjunction; Cl–clause; coord–coordinate conjunction; doaux–"do" verb; ger–gerund; havaux–"have" verb; int–interrogative; mod–modal verb; n–noun; NG–noun group; NP–noun phrase; ord–ordinal numeral; part–past participle; Ph–phrase; pred–predeterminer; PP–prepositional phrase; Pr–prepositional group; prep–preposition; prn–pronoun; qt–quantifier; S–sentence; v–verb; VG–verb group; VP–verb phrase.

(Note: A unit in parentheses is optional; * indicates a concatenation of one or more units; for multiple units enclosed by brackets, choose one unit.)

Grammar:

S → [(cj) Cl]*
Cl → Ph*
Ph → NP, PP, AdvP, VP
VP → VG, beaux AdjP, VG coord VG
AdjP → (adv) [adj, part] (coord (adv) [adj, part]), "*as* adj *as*" [NG, Cl]
PP → Pr NP, Pr int VG
NG → (pred) ([art, (qt) (ord) (card)]) (AdjP)* [n, ger]*, prn, int
VG → ([mod, doaux]) v, mod [be, have been] [part, ger], mod have part, beaux [part, ger], havaux part, havaux been [part, ger]
NP → ([we, you]) NG, ([that, int]) Cl (prep), NG AdjP, [more, less, -er] (NG) than [Cl, NG], (not) [ger, to v] ([NP, PP, Cl]), NG coord NG, as (many, few, adj a NG) as (NG, Cl)
Pr → prep (coord prep)
AdvP → Freq a NP, as adv as [Cl, adj], (Freq) [more, less] adv than Cl
Freq → once, twice, [qt, card] times

Other rules/constraints: 1. Cl must contain a NG and a VG (except for the main clause, which may begin with an imperative (unconjugated) VG and no NG); 2. sentential adverbs and PPs may appear adjacent to any

verb; 3. "not" may follow any modal or auxiliary verb, 4. "all" may follow any modal or auxiliary verb, after a plural subject; 5. each subject NG must agree with its VG in number.

Appendix B

The following shows the parse results for three groups of sentences (96 in all), which had formed the test sets for three previous articles in the literature (Dewar 1969; Thorne 1968; Milne 1986). A capital letter code (the output of our local parse algorithm) is placed after each corresponding sequence of words, and a slash is noted at points where the global parser predicted a major syntactic boundary. The codes are: N–noun group; V–verb group; P–prepositional phrase; I–infinitival phrase; A–adverbial phrase; S–personal pronoun; R–relative pronoun; G–gerund phrase; C–conjunction; X–auxiliary verb. Parsing mistakes are noted with numbers and are explained below.

GROUP 1

John N helped V Mary N.
John N helped V the girl N.
The boy N helped V the girl N.
Why R did X the chicken N cross V the road N?
Chew N gum V.[1]
We S are going V to London P.
Last week N / we S visited V John N.
It S is easy V / to make I a mistake N.
Anyone N can make V a mistake N.
How difficult N was X it S to find digs I?
He S observed V the girl N with the telescope P.
He S observed her V with the telescope P.
Which magazines N do X you S prefer V?
The cat N and C dog N play V.
Candy N is dandy V / but C liquor N is quicker V.
I S like V bathing beauties N.
Will X you S tell V John N / to bring back I the book N?
I S can give you V a rough estimate N.
Did X you S see V the house N / he S built V?
The film N / which R Punch N recommended V / was banned V.
The policeman N stopped V and C questioned him V.
He S likes V / reading G Shakespeare's play N / and C performing them N.
Take V an egg N and C beat it V.
The butler N did V / what R we S wanted him V to do I.
Where R have X you S and C your father N been V hiding G?

GROUP 2

She S visited him V yesterday N.
He S must have moved V.
The cat N adores V fish N.
This cat N adores V fish N.
Go V.
Mary N hates V my teasing N her S .
The boy N / who R kissed V the girl N / laughed uproariously V.
The boy N / who R the girl N kissed V / laughed uproariously V.
The boy N / the girl N kissed V / laughed uproariously V.
Fred N gave V the dog biscuits N.[2]
Fred N lost V the dog biscuits N.
Pick V a ripe banana N.
When R did X John N say V / he S would come V?
He S observed V the man N with the telescope P.
Power N corrupts V.
He S is waiting V.
Playing cards G intrigues me V.
Playing cards G intrigue me V.
I S dislike V playing cards G.
The rascal N / who R John N claimed V / committed V the crime N has escaped V.
Whose book N did X you S say you V wanted V?[3]
When R he S has fixed V dates N / he S will ring us V.
The queen's sister's husband N took V good photographs N.
A lawyer N / who R cheats V the clients N / he S sees V / deserves V censure N.
Has X the portrait N / they S bought V / disappeared V?
He S rolled V up the bright red carpet P.[4]
She S handed V John N / a pear N and C Mary N / an apple N.
The plants N / he S watered V and C tended V / flourished V.
Are X the elephant N and C the kangaroo N / he S adopted obeying him V?[5]

GROUP 3

The old N can get V in A / for C half N price V.[6]
The large student residence N blocks V my view N.
I S know V / that R boy N is bad V.[7]
I S know V / that R boys N are bad V.
What boy N did it V?
What boys N do V / is not V my business N.[8]
The trash N can be smelly V.
The trash can N was smelly V.
Which boy N wants V a fish N?
Which boys N want V fish N?
The river N / which R I S saw V / has V many fish N.
What boy N wants V a fish N ?
What boys want N is V fish N.[9]
What R blocks V the road N?
What blocks N are V in the road P?
What R climbs V trees N?
What climbs N did X you S do X ?
I S know V / that R boy N should do it V.[10]
I S know V / that R boys N should do it V.

That R deer N ate V everything N in my garden P / surprised me V.
That deer ate everything N in my garden P last A night V.[11]
I S know V / that hit Mary N.[12]
I S know V / that N will be true V.
I S know V / that R boys N are mean V.
I S know V / that R Tom N will hit V Mary N.
I S told V the girl N / that R I S liked V the story N.
I S told V the girl N / whom R I S liked V the story N.[13]
I S told V the girl N / the story N that R I S liked V.
Have X the students N take V the exam N.
Have X the students N taken V the exam N?
The soup pot cover handle screw N is red V.
The soup pot N cover V handles N screw tightly V.[14]
The soup pot cover handle N screws tightly V.
The soup pot cover handle screws N are red V.
Which years N do X you S have V costs N figures N for P?
Do X you S have V a count of the number of sales V requests N and C the number of requests N filled V?
The trash N can N was taken out V.
The trash N can be taken out V.
The paper will N was destroyed V.
The paper N will be destroyed V.
Let X the paper N will N be read V.
Will X the paper N can be re-used V?[15]

Notes

1. Noun-verb confusion.
*2. Separating the direct and indirect objects is difficult here.
3. The second "you" is treated as a direct object, rather than as a subject.
4. The adverb "up" is mislabeled as a preposition.
5. "Obeying him" is not syntactically linked to "are".
6. The sequence "in for" causes "for" to be treated as a conjunction.
*7. Without a semantic component, "that" is mislabeled as a relative pronoun.
8. "What boys" is treated as a unit.
*9. "What boys want" is treated as a unit.
*10. Same error as in (7).
11. The first four words are grouped as a noun phrase.
12. "That" is mislabeled as a demonstrative.
*13. The end of the embedded clause is not assigned a major boundary.
14. "Cover" is mislabeled as a verb.
15. "Can" is mislabeled as a verb.

The mistakes labeled with an asterisk do not cause any intonation errors, and most of the other sentences which caused syntactic problems would have only slight intonational variations as a result of the mistakes.