An Experimental Discourse-Neutral Prosodic Phrasing System for Mandarin Chinese

Gina-Anne Levow
Massachusetts Institute of Technology
Cambridge, MA 02139
gina@ai.mit.edu

Abstract Normal human speech has a clear intonational and rhythmic character. This is true of many Pacific Rim languages and plays a particularly crucial role in the many tone languages of the region, such as Thai and Chinese. However, most computer speech systems fail to utilize prosody for disambiguation or increased naturalness. In this paper, we examine the contrasts in prosodic behavior across languages by implementing a prosodic phrasing system for Mandarin Chinese and comparing its output phrasings to those produced by actors in the running dialogue of a real Chinese film. Drawing on an experimental English-language prosodic system developed by Bachenko *et al.* to show how to produce more natural computer speech, we have designed a discourse-neutral system for Chinese to test the cross-linguistic applicability of the well-known interaction among syntactic structure, constituent length, and prosodic structure to computationally predict the prosodic phrasing of a sentence.

The experiment then compared these phrasings against a corpus of 116 natural utterances from a Chinese film. Revisions to the system tested the relevance of additional syntactic information. This additional information about universal relations such as syntactic constituency and domain of modification substantially increased phrasing accuracy, bringing the number of correctly phrased sentences from 39.65% to 70%.

The success of these new algorithms supports the assumption of a close relationship between syntactic and prosodic structure across languages. However, the improvements gained by increased use of sentence and verb phrase constituent information indicate that English and Chinese differ in the types of syntactic information that are most needed. The unresolved errors of the system also demonstrate the insufficiency of a purely syntactic model. A fully adequate system for prosodic phrasing will need to exploit additional semantic and pragmatic constraints.

1 Introduction

Over the past twenty years, there have been significant improvements in computer speech systems. Speech recognition systems can now attain high accuracy rates for speaker-independent continuous speech. Speech and text systems are used in a wide variety of applications, from telephone directory assistance to voice-mail. However, in spite of these dramatic changes in utility and accuracy, computer speech systems still do not approach the prosodic output capability of human language users.

This paper represents a departure from previous research in three main respects. First and foremost, it examines the cross-linguistic applicability of approaches to predicting utterance prosodic phrasing based on the relation among syntactic structure, sentence length, and prosodic structure. In particular, the primary language of study is Mandarin Chinese, in contrast to analyses of American English. Secondly, the experimental corpus used for development and testing of the prosodic phrasing system and the parser is a long fragment of the running dialog of a real Mandarin Chinese film produced in the 1940's. The audio component of this film as well as the judgments of native speakers provided the metric against which to judge the accuracy of the phrasing system. Finally, the increased use of universal structural relations such as constituency and domain of modification led to a substantial improvement in the phrasing system's accuracy.

One domain in which the performance gap between human and machine is particularly evident is in the handling of prosody and intonation. Prosody includes the stress, timing and pitch features of segments as well as supra-segmental elements such as words, phrases, and sentences. This aspect of speech subsumes areas from syllable stress to the intonational pitch contour that marks the focused element in languages such as Bengali.

Even though prosodic features represent some of the most noticeable components of speech, prosodic information is little-used in current computer systems. In most speech recognition systems, the prosodic information is stripped away from the signal prior to processing and thus plays no role. In speech generation systems, only rudimentary features such as the falling contour associated with a declarative utterance and the rising contour of an interrogative utterance are generally incorporated.

Given the successes of speech systems, why should we be concerned with the absence of prosodic infor-

mation? There are a number of compelling reasons to explore and expand the use of prosody in speech systems. One main way in which prosody is exploited by human speakers is for utterance disambiguation. Tone languages differentiate between word meanings on the basis of lexical tones, languages with syllabic stress differentiate word meanings on this basis, and, further, prosodic phrasing - the placement of phrase boundaries and pauses of varying saliences which is the focus of the current study - can be used to differentiate between ambiguous structural interpretations, as in the case of 'garden path' sentences, such as "The horse raced past the barn fell." Prosody also contributes to the naturalness of speech; human beings do not speak in a monotone with no pauses for breath. Finally, semantic and pragmatic features such as emphasis and contrast are expressed prosodically through the use of special intonations.

2 Prosodic Phrasing System

This study implemented a computer system to computationally evaluate sentence structure and length as they relate cross-linguistically to the prosodic phrasing of an utterance. We implemented a prosodic phrasing system for Mandarin Chinese based on concepts set forth in a paper by Bachenko *et al.* (1990) on discourse-neutral prosodic phrasing in English. The system takes the *pinyin* Romanization of a Mandarin sentence string and produces a prosodic phrase structure tree and an annotated sentence string which indicate the predicted locations and relative saliences of the prosodic phrase boundaries of the utterance. While Bachenko's initial system achieved fully correct sentence phrasings for only 39.65% of the Mandarin test utterances, the final version of the system which had been modified to address faults in the initial approach achieved a success rate of 70%.

2.1 Features of Study

Let us now turn to an examination of the discourse-neutral prosodic phrasing system developed for Mandarin Chinese. This study raised the question of whether the system developed by Bachenko applied cross-linguistically and introduced some novel design challenges. First, the primary language of study is Mandarin Chinese. Mandarin has a full system of lexical tones as well as sentence intonation. There are four canonical lexical tones - 1) high - level, 2) high - rising, 3) low - falling rising, and 4) high - falling. There is also a fifth, neutral tone that occurs on unstressed syllables. A common example of lexical tone is the word 'ma', which has the following meanings with the given tones: first - mother, second - hemp, third - horse, fourth - scold, neutral - yes/no question particle. Recent studies of sentence-level intonation by Shen (1989) indicate that declarative utterances are marked by lowering of average pitch while interrogatives exhibit higher average pitch. These sentence-level intonational patterns map onto the basic contours prescribed by lexical tones and may be viewed as overlaying the lexical tone structure. Mandarin Chinese also has a fairly freely-ordered phrase structure. Verb phrase elements and modifiers such as objects and locatives can appear in a variety of pre- and post- verbal positions. As a result, it exhibits both subject-verb-object and subject-object-verb surface structures. This behavior is in marked contrast to the rigid S-V-O ordering imposed by English. These features provide an interesting context for the study of prosodic phrasing and provide a basis for comparison with English phrasing systems.

The corpus of sentences used for the design, testing, and analysis of the prosodic phraser was drawn from the audio component of an actual Chinese film, *Ai Le Zhongnian, The Joys and Sorrows of Middle Age*, as prepared by Liu and Wickeri. The utterances thus came from a fragment of a real running dialog, and are representative of colloquial speech. The film itself is used in a number of programs for instruction in Mandarin and exhibits natural speech with clear production. A suite of approximately 120 sentences was used for analysis, while a significantly larger set was used to develop the parser and algorithm.

2.2 Phrasing Challenges

A number of issues make the development of prosodic phrasing systems challenging and interesting. On one hand, languages themselves often provide useful cues for prosodic phrasing. One of the clearest examples of such cues is the English punctuation system, which can clearly delineate clause boundaries, set off initial phrases and clauses, and mark off appositive elements. It would seem simple enough to just extract the phrasing from the punctuated utterances. However, the problem of prosodic phrasing cannot be solved so easily. While punctuation in English can provide useful cues that are exploited by an number of text-to-speech systems including that by Bachenko, many languages do not employ such a well-delineated system of markings. Thus, helpful cues in one language may mislead or be absent in many other languages.

Some of the issues raised by prosodic phrasing systems may also contribute to work on wider areas of language processing. The prosodic features of a sentence may determine components of its semantic meaning and its pragmatic use of intonation to designate emphatic and contrastive stress and focus.

A better understanding of the extent to which syntax influences prosodic phrasing can help to isolate the roles which semantics and pragmatics play in the intonational structure of an utterance.

Some of the difficulties encountered in discourse analysis have clear parallels in prosodic analysis. Cues, such as the use of certain identified 'cue words', often play a role in discourse processing. Phrases like 'on the other hand' often occur at a return to a previous discourse segment. However, as in the case of punctuation, these hints are not necessarily present in all situations in which a pop of the context stack is required. Perhaps the exploration of the mismatches which occur between cues and prosodic structure and their avoidance may point out profitable lines of inquiry for discourse systems.

In order to meet these challenges in a less language-specific context, we made use of relations that seem to be universal and hold across many different languages. The relations between elements and constituents that arise from syntactic structure provide one such set of universals. The notions of noun and verb classes and associated phrasal units appear throughout the world. Also, the notion of scope or domain of modification will prove to be another useful, broad-based concept.

2.2 System Overview

Next, let us take an overview of the input-output behavior of the implemented prosodic phrasing system. The system takes an unpunctuated sentence string in which the basic elements are *ci*, word-level meaning units, Romanized according to the widely-used *pinyin* transliteration system. A copy of the string is then passed to a parser developed for this system. The parse tree and the original string are then both passed to the core of the prosodic phrasing system. The output appears in two formats. One is a prosodic phrase tree in which the leaf nodes correspond to words and the root and internal nodes indicate the relative saliences of the boundaries that occur between the two child phrases of the current node. A simpler annotated sentence string is also produced in which the boundaries between prosodic phrases are indicated by either !, |, or ||, where the symbols are in order of increasing salience. A diagram of this system structure appears in Fig. 1.

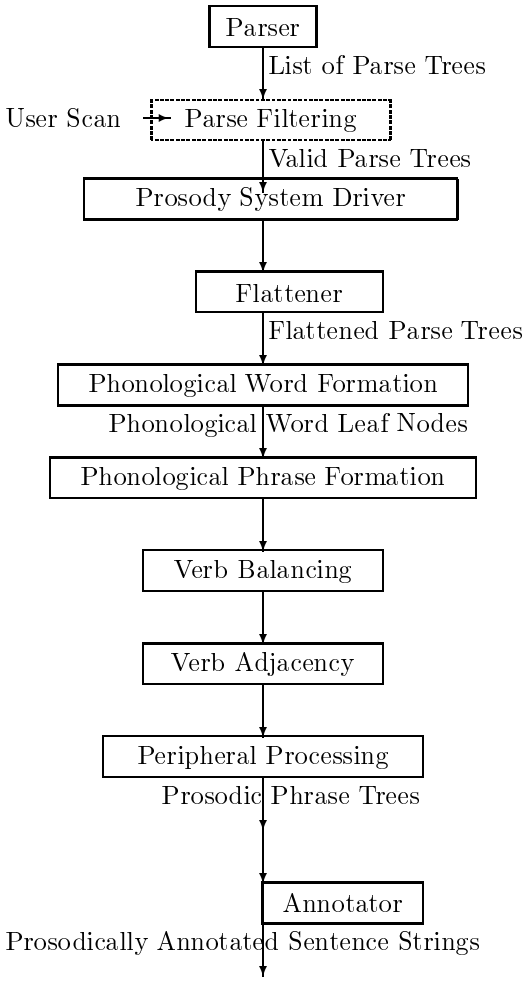


Figure 1: Prosodic Phrasing System Block Diagram

2.3 System Design

The prosodic phrasing system underwent an iterative development process at each stage of which the errors of the current implementation were analyzed and changes were made to address the sources of the errors. There were three main development phases:

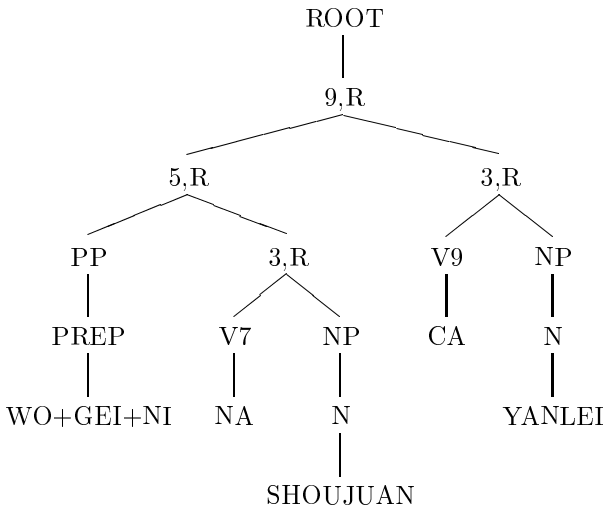


Figure 2: Prosodic Phrasing Tree for "Wo gei ni | na shoujuan || ca yanlei" with annotation

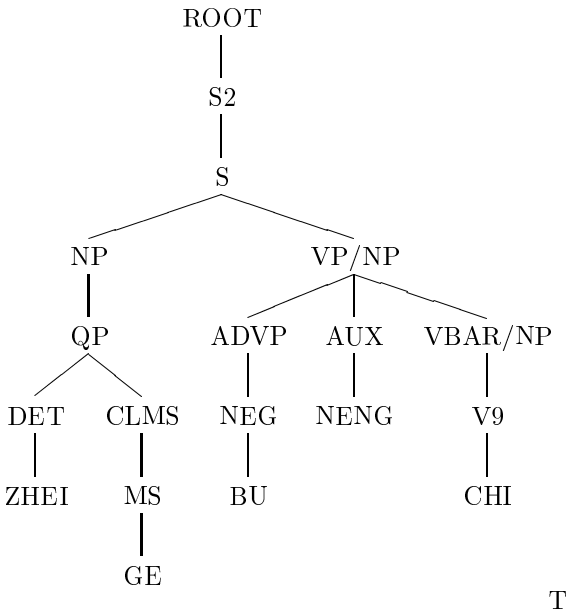
original, revision 1, and revision 2. We will examine each in more detail below.

2.3.1 Original System Development

The original phase of development involved the adaptation of the English prosodic phrasing system described by Bachenko *et al.* (1990) to handle Mandarin Chinese. The first step was the design of the parser and its associated lexicon. The parser as implemented was an Earley-style parser, working bottom-up with top-down filtering. The parser used 176 context-free phrase structure rules, many of which were included to make explicit the structure of verbal complements and adjuncts and to handle the syntactic behaviors of different classes of nouns and time words.

Following Bachenko *et al.* (1990), a routine was included which 'flattened' a syntactic structure tree. Bachenko argued, based on sentences such as "I believe California sales are still off by 20%" in which no phrase boundary occurs between 'believe', the verb of the matrix sentence, and 'California sales', the subject of a clausal argument, that no distinction should be made between arguments and adjuncts of verbs or between constituents of different clausal elements. Thus, after parsing, all sentence and verb phrase nodes are removed from the parse tree producing a 'flattened' structure. A phrase structure tree and its flattened counterpart are shown in Figures 3 and 4, respectively, for "Zhei ge bu neng chi",

"This, (you) can't eat."



This (you) not can eat.

Figure 3: Initial Parse Tree for "Zhei ge bu neng chi"

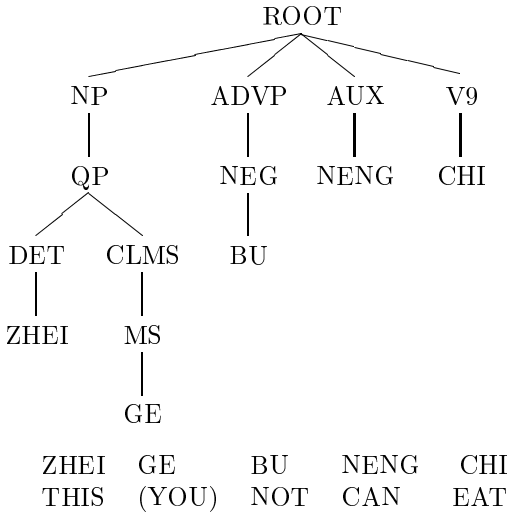
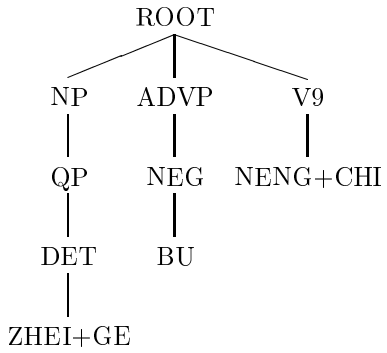


Figure 4: Flattened Parse Tree for "Zhei ge bu neng chi"

The remainder of the phrasing system then uses this flattened parse tree and the original input string to produce a prosodic phrase structure tree. The first

stage of this process involves the formation of phonological words. These words are the basic units of prosodic analysis. They are analogous but not identical to lexicographic words. The fundamental idea behind their use is that not all words should be treated in the same manner for phrasing purposes. Function words such as pronouns, auxiliaries, prepositions, and articles are often unstressed or destressed in speech in contrast with content words such as nouns, verbs, adjectives, and adverbs. For example, consider the contraction of 'going to' to 'gonna' and the changes from the long e in isolation of 'the' to a schwa in 'the dog.' Thus, following Bachenko *et al.*, we adopt a form of Grosjean and Gee's (1987) definition of phonological words as one or more function words linked with an associated content word. Each phonological word is treated as a single unit for the remainder of prosodic analysis, as in Figure 5.



This (you) not can eat.

Figure 5: Tree of Phonological Words

The next stage involves the formation of phonological phrases from phonological words. This process simply traverses the sentence from left to right and groups all of the words up to the next major syntactic head, a noun or a verb. The goal in this step is to link all pre-head modifiers with the appropriate head. In the case of a noun phrase of the form Det+ADJ ADJ N, after phonological word formation , these elements would combine to form a noun phonological phrase. In the case of a verb, all the modals and auxiliaries words link to the verb itself, as seen in Figure 6.

Now that the phrases of the prosodic structure have been formed we can proceed to the main goal of the system, the determination of the placement and saliences of the boundaries between phrases. The core process is termed verb-balancing. In Bachenko

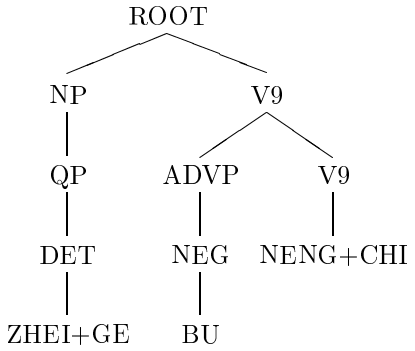


Figure 6: Tree of Phonological Phrases

et al. (1990)'s analysis the verb is considered to be the central or pivotal element in semantic, syntactic, and prosodic structure. Further, since length between prosodic boundaries has been shown to be a clear predictor of the location of the next boundary [Hirschberg & Wang 1990], it is considered desirable for prosodic phrases to be close to the same length. Length, or weight, in this context is defined to be the number of phonological words in the phrase. Verb-balancing operates as follows: if L[X] + L[V] > L[Y], then X(VY), otherwise (XV)Y, where V is the verb phonological phrase, X is the pre-verbal constituent, Y is the post-verbal constituent, and either X or Y may be nil. This statement means that if the combined weights of the pre-verbal and verbal phrases are greater than that of the post-verbal element, then group the verb with the post-verbal phrase; otherwise, group it with the pre-verbal phrase. This activity corresponds to balancing the weights of prosodic phrases around the verb.

The next stage of operation, which is very closely associated with verb-balancing, is called verb-adjacency by Bachenko *et al.* (1990). This step simply groups the excluded element from the verb-balancing operation with the node formed in the preceding step. Between the two units of each cluster, a phrase boundary is assigned with weight equal to the sum of the weights of the paired phrases plus one. A threshold was set at a weight of five; boundaries below this weight which lacked internal structure were not considered to be salient. (Figure 7)

The remaining constituents in the utterance, those not adjacent to a verb, were handled by 'peripheral processing.' These phonological phrases were linked one at a time together up to the main node and one

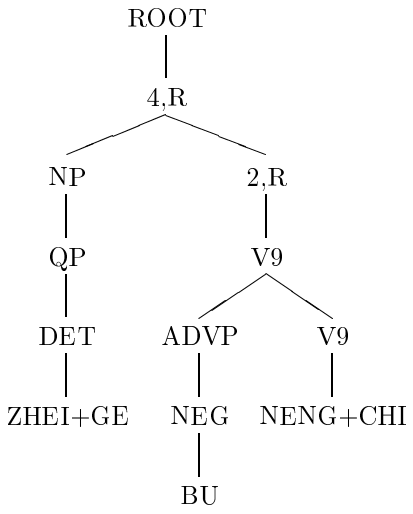


Figure 7: Tree after Balancing and Adjacency

at a time to the main node if the elements occurred to the right. A phrase boundary of added salience would only be postulated if the rightward constituent were more than one word in length. This limitation was intended to prevent long sequences of post-verbal modifiers from being set off by boundaries of rapidly increasing salience. This process produces the final output prosodic phrase tree with indicated boundary saliences. The annotated sentence string is then produced by reading the relative saliences from the tree.

2.3.2 Special Modifications for Chinese

Each of the steps described above was modified to handle the new features introduced by the Mandarin Chinese domain. In the case of phonological words, the behavior of new types of function words in Chinese needed to be analyzed to determine correct adjoining behavior, as with the *de*-relativizer which links to the left toward the modifier. The adjoining of function types with English parallels also required consideration to avoid creation of false similarities. The formation of phonological phrases was also affected by the introduction of new syntactic phrase types as well as different structures for previously encountered categories, as in the case of strictly head-final noun phrases and pre-verbal modifiers. While the basic approaches in verb-balancing and verb-adjacency were quickly adapted to the Chinese domain, peripheral phrase handling had to be expanded to cope with a large number of pre-verbal and sentence-initial peripheral constituents.

Once all of these basic modifications had been made and tested for operation, the system was run on a suite of utterances from the transcript of the film, and the results were compared to the utterances in the audio component of the original film and to the judgments of linguistically naive native speakers, who recorded their judgments by annotating a textual transcription of the utterances with their preferences for boundary location and salience. The system phrased 39.65% of the utterances in full agreement with the recorded utterances. The remainder of the phrasings each contained at least one error. The misphrasings were analyzed to determine the underlying causes of the errors. In some cases, more than one error type contributed to the misphrasing of an utterance; in these instances, multiple error classes were recorded. A detailed schematic breakdown of the number and types of misphrasings appears in Figure 8. The most common sources were misadjoined pronouns, misaligned syntactic structures, and misscoped adverbs and time phrases. Other errors, such as those due to mishandled noun phrases or assignment of emphatic or contrastive stress, accounted for the remainder of the instances.

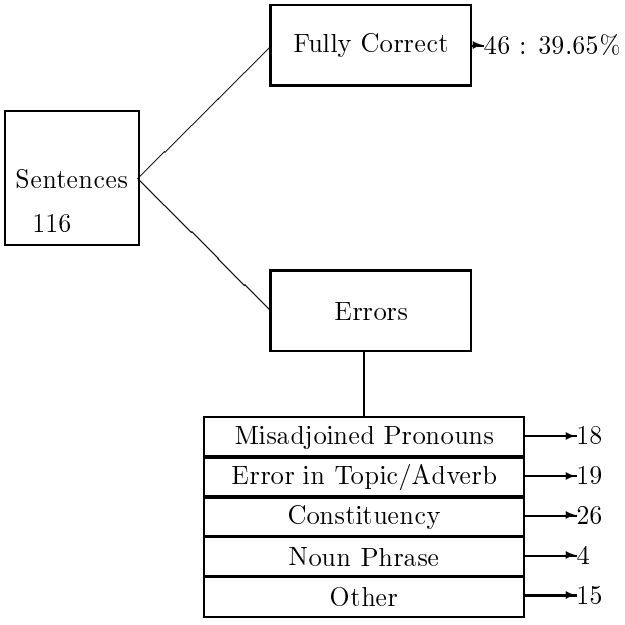


Figure 8: Error Breakdown for Base Algorithm

The root of the majority of the mislabelings can thus be traced to difficulties related to syntactic con-

stituency, in particular, to the loss of constituent information due to the flattening of syntactic structure to remove verb phrase and sentence-level nodes. Thus, although the parser would obtain a tree which correctly disambiguated structures for phrasing, the flattened version proved insufficient for the phrasing task.

2.3.3 First System Revision

The next stage of the iterative development process then involved the retention of this syntactic information into the core of the prosodic phrasing system. Since a large proportion of the errors appeared as misadjoined pronouns, we tried to prevent this mislinking by applying the full syntax tree to the formation of phonological words. It was reasoned that this inclusion would allow the differentiation of

wei le ni , baba zuo shi

and

wei le ni baba zuo shi

where the first utterance means "It is for you that Father works" and the second means "It is for your father that (you) work." Yet another means "For your father to work." The confusion in the first case is caused by the need to determine whether 'ni' (you) should be grouped as the object of the preposition 'wei le' or as the possessive modifier of 'baba', in the absence of distinct case marking for 'your.'

2.3.4 Final System Revision There was, as expected, a substantial improvement in the action of the prosodic phraser as many of the misadjoinings were removed. However, a number of errors related to syntactic constituency persisted. One common error configuration involved the association of material between two verbs. Since many modifiers may appear in pre- or post-verbal position in Chinese, in the absence of clause or verb phrase constituency information the algorithm can not determine whether elements between verbs should be associated with the preceding or following verb during verb balancing and verb adjacency. Phrases with the locative preposition 'zai', for instance, may come before or after the verb. In order to try to resolve these difficulties, the second revision made use of sentence and verb phrase constituency for linking through phonological phrase formation and verb balancing.

The test suite was run on both revised versions of the prosodic phrasing system. Again the results were broken down into error classes according to the source of the misphrasing. Now over 70% of the sentences were phrased with full accuracy. The remaining utterances revealed that the errors arose from issues surrounding the scope of the negative particles, the handling of certain pre-verbal structures, idiom analysis, complex noun phrases, and emphatic usages. The difficulties associated with constituent mismatches have mostly been resolved. The final error breakdown may be seen in figure 9 below.

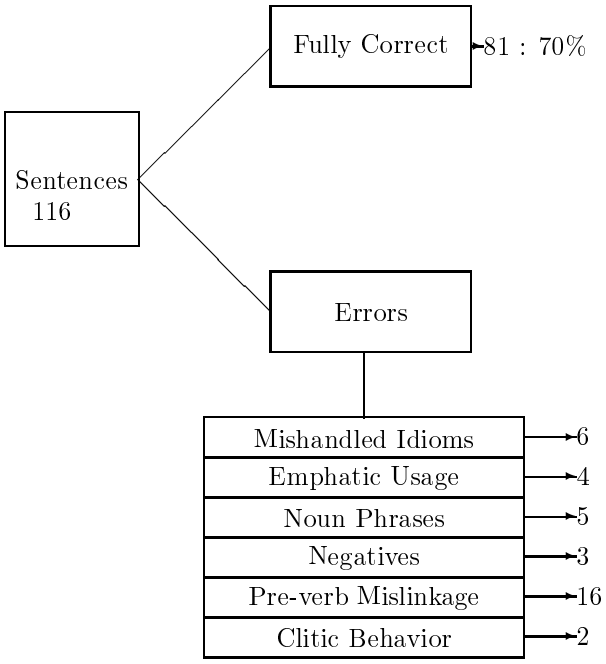


Figure 9: Error Breakdown for System Revision 2

Version	*null bound*		*salient bound*	
	acc	succ	acc	succ
Original	81.4%	91.4%	73.9%	49.7%
Revised 1	89.9%	92.4%	74.8%	65.3%
Revised 2	92.4%	95%	84%	80%
Null	70.8%	100%	0%	0%

Table 1: Accuracy and Success Rates for Null and Salient Boundary Labeling Compared to Human Judgments

3 System Analysis

3.1 Challenges to Underlying Assumptions

There are a number of underlying assumptions made in the prosodic phrasing system designed by

Version	*Overall Averages*		*Ratings*
	accuracy	success	success
Original	77.64%	70.56%	79.26%
Revised 1	83.4%	78.85%	84.49%
Revised 2	88.18%	87.5%	90.73%
Null	35.4%	50%	70.8%

Table 2: Average Accuracy and Success Rates and Overall Success

Bachenko *et al.* which were called into question as the system was adapted to a new linguistic context. It is the ability of cross-linguistic analysis to highlight these implicit assumptions which makes such methods valuable. In the following paragraphs we will examine some of these beliefs and the problems that they raised for Mandarin Chinese phrasing.

One of the most evident assumptions is that of explicit, precise punctuation. Bachenko's algorithm makes direct use of sentence and phrase punctuation in the placement of prosodic phrase boundaries. The basic Chinese language lacks not only punctuation, but also interword spacing. While some aspects of English punctuation have found their way into use in parts of China, the rules for their use are not hard and fast, and indeed they are used rather minimally. Thus one can not rely upon punctuation to guide prosodic phrasing in Mandarin Chinese.

Another assumption made by Bachenko *et al.* which was never incorporated into the Chinese system was a so-called 'left corner constraint'. This rule placed a phrase boundary at the left edge of a conjoined phrase, to the left of the coordinating morpheme itself. In English, this morpheme is a good predictor of a phrase boundary; however, coordinating morphemes are not so prevalent in listing constructs in Chinese. The same phrase break occurs, but there is often no explicit morpheme to mark the split between elements of a pair or a list.

The impact of the next several issues to be discussed became apparent only after their inclusion in the Chinese prosodic phraser caused errors. The first of these in order of processing involves the formation of phonological words. Pronouns form one of the most frequently encountered groups of function words. Their adjoining behavior in Bachenko's approach is characterized as follows: subject (nominative) pronouns should adjoin toward their verb and object pronouns (preposition or verb) should adjoin toward the left. In English, the determination of type of pronoun may be made quite easily on the basis of case. A subject pronoun will have nominative case, whereas an object pronoun will get accusative case, as in the contrast between 'I' and 'me.' This simple solution is not available in Mandarin Chinese. The form of a pronoun does not differ based on use; case is not overtly marked. The situation is further complicated by an additional use of pronouns as modifiers of nouns to indicate possession. The role played by the pronoun, however, may be determined by the syntactic structure of the sentence, by the government and subcategorization characteristics of the verb. Thus, these general syntactic consideration can often resolve pronominal ambiguity.

The formation of phonological phrases also proved to be too simple for some of the constructions present in Chinese. In English, major syntactic heads, nouns and verbs, are generally preceded by some small number of pre-head modifiers while other complements and adjuncts follow. In Chinese, in contrast, noun phrases are strictly head-final, all modifiers precede the head noun, and long strings of time words, locatives, pre-complementizers, and fronted elements may appear before the verb. These clusters are often much too large to be treated as a single phonological phrase, but the base algorithm offers no recourse.

As discussed earlier, one of the basic tenets of the verb-balancing approach to phrase boundary placement is the centrality of the verb. In English, a strict subject-verb-object language, the location of the verb in syntax supports this treatment very strongly. However, the syntactic structure of Mandarin Chinese has been characterized both as subject-verb-object and as subject-object-verb. Both forms of surface structure occur in the utterances examined. Thus the appropriateness of this approach is not as apparent. It was hoped that experimentation would indicate the extent to which verb-balancing failed or succeeded in predicting the prosodic phrasing of Mandarin Chinese.

The inclusion of syntactic constituency information which in Bachenko's algorithm was removed by the flattening operation on the parse tree allowed the determination of relationships needed for phrasing. This case was illustrated in the discussion of phonological word formation with pronouns. Clausal constituency also plays a major role in determining phonological phrase formation in the case of adverbs which modify the sentence as a whole, in contrast to those which modify only the adjacent constituent or group. Verb-balancing can also be affected by the association of verb-medial elements with the preceding or following verb. This increased use of syntactic structure allowed a substantial improvement in accuracy of phrasing along two metrics, accuracy

and success, analyzed for null and salient boundaries separately and over both boundary types. 'Accuracy' represents the percentage of labelings of a possible boundary location, any position between words, which were correct; 'success' represents the percentage of boundaries of a given type which were labeled correctly. These measures were used in conjunction to identify false positives and false negatives. The improvements of the two revisions over the original and a baseline in which all possible boundary locations are marked null or boundariless are reflected in the Tables 1 and 2.

3.2 Unresolved Issues and Future Work

In spite of the improvements in prosodic phrasing achieved by the modifications to the prosodic phrasing system, there are still a number of issues which remain unresolved. First, the effect of S-O-V phrase structure on the success of the algorithm was difficult to determine. There did not seem to be any significant errors directly attributable to phrase ordering. The verb-balancing operation would indicate some phrasing for the immediate vicinity of the verb, and peripheral operations produced reasonable boundary positions for the rest of the utterance. However, this result should not be taken to indicate a full endorsement of this approach. It can be argued that the corpus did not contain utterances which presented a sufficient challenge to verb balancing. While approximately one-fourth of the test utterances had no post-verbal material, most of the utterances were short. Many of them fell below the five phonological word threshold for phrase boundary salience, and the remainder had only a single salient boundary which naturally was posited to fall before the verb. It would be interesting to more fully evaluate the notion of verb centrality by testing the algorithm on a true S-O-V language such as Japanese.

The handling of complex noun phrases still poses a challenge. In Chinese, noun phrases may be composed of any number of adjectival, possessive, verbal, or relative clause modifiers, each optionally followed by the *de* -relativizer, in front of the head noun. A long concatenation of such modifiers should, intuitively, be separated by one or more phrase boundaries. However, the positioning of these boundaries must depend on the relations of modification among the elements, which may impose some structure or hierarchy. These relations can not be determined without some knowledge of the legal semantic relations and the speaker's intended meaning.

A different set of semantically-related difficulties arise in the phrasing of utterances which include idioms. A simple example of such idioms is the common phrase 'mei you yisi', which literally means 'does not have meaning' but has the idiomatic sense of 'isn't interesting.' Syntactically, the phrase is of the form 'Neg V N', but due to the desired idiomatic treatment one does not wish to assign a salient phrase boundary between 'you' and 'yisi' regardless of the length of the pre-verbal constituent phrase. Thus, one can not necessarily make the same phrasing judgments based on the syntactic structure of an idiom that one might wish to make for any other phrase with similar form. The phrasing is dependent on the presence of an idiomatic interpretation.

Finally, there are issues which seem to be related to the pragmatic features of the utterance. For instance, certain words carry emphatic stress and thus affect the prosodic structure of the utterance. Words receiving stress in Mandarin Chinese experience a lengthening of prosodic features such as length and duration as well as an exaggeration of the differences in pitch contour. It stands to reason that overall prosodic phrasing would be affected by such drastic changes. Words which exhibit such behavior in Chinese are *cai* and *tai*, corresponding to 'only/until', and 'too' in English. It has been noted by some researchers that 'only' and 'until' have similar effects in English as well.

The remaining instances of pragmatic stress derive from contrast and emphasis specified by the larger discourse context. One such instance occurs when the protagonist reports on his attempt to act as a go-between for a friend and says "Ta daying gen WO jiehun", "She agreed to marry ME." Such externally-motivated intonational patterns naturally affect the prosodic phrasing in ways which can not be predicted purely from the structure of the utterance. Since the current study specifically involved only discourse-neutral prosodic phrasing, these phenomena are, strictly speaking, outside the range of the study. However, they may indicate a new direction for exploration of prosodic behavior in the broader context of discourse effects.

4 Conclusions

The development of an experimental prosodic phrasing system for Mandarin Chinese allowed us to computationally analyze the relationship among phrase length, sentence structure, and prosodic realization. The study indicated that syntactic structure plays a significant role in determining the prosody of an utterance. Errors encountered by the initial implementation of the system highlighted the importance of syntactic constituency and domain of modification as well as the implicit and sometimes erroneous assumptions caused by reliance on cues to prosodic phrasing

which are not available cross-linguistically. The increased use of universal structural relations, such as constituency and modification scope, allowed the system to achieve a final overall success rate of more than 90% for prosodic phrase boundary marking.

One can draw a potentially useful analogy between prosodic phrasing and discourse analysis. As the study noted, there are many useful cues to prosodic phrasing which occur in some languages, such as punctuation and case marking. A strong parallel exists in the realm of discourse in which cue words and phrases, such as 'anyway' and 'on the other hand', often play a major role in discourse systems for segmentation and analysis. Most researchers will agree that such cues are felicitous but certainly not obligatory for change of focus or manipulation of discourse representations. The results of the prosodic study indicated that universal relations of the underlying syntactic structure correlated with the observed prosodic behavior even in the absence of overt cues. Perhaps closer examination could lead to the discovery of similar universal relations in discourse that could illuminate its structure when no cue words appear.[1]

References

[1] Steven P. Abney. Parsing by chunks. In R. C. Berwick et al., editors, *Principle-Based Parsing: Computation and Psycholinguistics*, pages 257–278. Kluwer Academic Publishers, 1991.

[2] J. Bachenko and E. Fitzpatrick. A computational grammar of discourse-neutral prosodic phrasing in English. *Computational Linguistics*, 16(3):155–170, September 1990.

[3] John Bear and Patti Price. Prosody, syntax, and parsing. In *28th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference*, 1990.

[4] Yuan-Ren Chao. *A Grammar of Spoken Chinese*. University of California Press, 1968.

[5] Chauncy C. Chu. The passive construction: Chinese and English. *Journal of Chinese Linguistics*, 3:437–470, 1973.

[6] Gerald Gazdar. Unbounded dependencies and coordinate structure. *Linguistic Inquiry*, 12(2), 1981.

[1]This paper describes research conducted at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. This project was supported by an graduate fellowship from the Office of Naval Research and the Vinton Hayes Fellowship to Prof. R. C. Berwick.

[7] J. P. Gee and J. Grosjean. Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15:411–458.

[8] Francois Grosjean and J. P. Gee. Prosodic structure and speech recognition. *Cognition*, 25:135–155, 1987.

[9] Anne Yue Hashimoto. Mandarin syntactic structures. *Unicorn*, 8:1–146, 1971.

[10] Cheng-Teh James Huang. *Logical Relations in Chinese and the Theory of Grammar*. PhD thesis, Massachusetts Institute of Technology, May 1982.

[11] Lin-Shan Lee, Lee-Feng Chien, Long-Ji Lin, Cheng-Teh James Huang, and K. J. Chen. An efficient natural language processing system specially designed for the Chinese language, 1991. Work in Progress.

[12] Charles Li and Sandra Thompson. *Mandarin Chinese: A Functional Reference Grammar*. 1981.

[13] Yen-Hui Audrey Li. *Order and Constituency in Mandarin Chinese*, volume 19 of *Studies in Natural Language and Linguistic Theory*. Kluwer Academic Publishers, 1990.

[14] Charles A. Liu and Janice Wickeri. *A Study Manual for "The Joys and Sorrows of Middle Age"*, volume 2 of *Chinese Film Series*. Princeton University, 1978.

[15] Jerrold M. Sadock. *Autolexical Syntax: A Theory of Parallel Grammatical Representations*. Studies in Contemporary Linguistics. University of Chicago Press, 1991.

[16] Hong San. *Ai Le Zhongnian*. Wenhua Movie Production Company, 1947. Film.

[17] Xiao-Nan Susan Shen. *The Prosody of Mandarin Chinese*, volume 118 of *University of California Publications in Linguistics*. University of California Press, 1989.

[18] Chi-Lin Shih. Tone and intonation in Mandarin Chinese. Working Paper 3, Cornell University, June 1988.

[19] Mark Steedman. Structure and intonation in spoken language understanding. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, 1990.

[20] Michelle Q. Wang and Julia Hirschberg. Automatic classification of intonational phrase boundaries. *Computer Speech and Language*, 6(2):175–196, April 1992.