PROSODICAL LABELING FOR MANDARIN

Hana Třísková

Academy of Sciences of the Czech Republic, Oriental Institute, Prague

David Sehnal

Charles University, Institute of East Asian Studies, Prague

ABSTRACT

The present paper describes the software PALM, designed to grasp and analyze the rhythm of Mandarin utterances. The functions of the software were tested on a small database consisting of 23 sentences recorded in slow tempo and in fast tempo. As a first step, the utterances were prosodically transcribed. The transcription captures stress features and horizontal segmentation. Theoretical fundaments of prosodic transcription are outlined (a simplified version of prosodic transcription is proposed to be used in teaching Mandarin as a second language). Transcribed utterances were broken into entries corresponding to syntactic words, then labeled for various features (both prosodical and grammatical). Query function allows retrieval of the instances - either words, or syllables - sharing various combinations of features (for words: number of syllables, syntactic function, stress pattern etc.; for syllables: level of stress, tonality etc.). Count function allows statistical processing of the search results. PALM was designed as a tool for finding links between rhythmical structure of the Mandarin utterance, its grammatical structure and speech tempo.

1. INTRODUCTION

The use of large, computerized bodies of text for linguistic description and analysis has been one of the most rapidly developing fields of activity in the study of language. Corpora have became well-established as a resource for language research. The earliest available computer corpora were corpora of

written language. Written language materials have continued to be by far the most abundant and readily available, because they are easy to collect and computerize. As G. Cook pointed out, "*Linguistics has been guilty of an Orwellian doublespeak: professing the primacy of speech, practising the primacy of writing*."[1] In the past decade, however, spoken language corpora have been given increased attention (there still seems to be a certain divergence between those working in speech and those in written language).

Language corpora, spoken and written, have been developed for a broad variety of linguistic studies. The results achieved in basic research find a wide range of practical applications. One of the reasons is the increasing recognition of corpus linguistics in many fields. In some cases, the corpora were designed for a specific purpose. For instance, the COBUILD collection (University of Birmingham, started in 1980) was designed specifically to provide a basis for the development of English dictionaries and learning materials. Other corpora were planned without an immediate project in mind, intended to be open to a wide range of future uses.

Transcribing spoken language already has a history of several decades. As early as in 1959, the SEU (Survey of English Usage) corpus, using manual methods and typewritten slips, was conceived. In 1975 it was decided to make this corpus available in a machine-readable form. The result became known as London-Lund corpus (LLC), which is the oldest and best known of the computerized spoken corpora. The MARSEC corpus (Machine Readable Spoken English Corpus, Roach et al. 1993) was derived from the SEC Lancaster/IBM Spoken English Corpus (compiled 1984-87). One of the largest corpora, the British National Corpus (BNC, started 1991, completed 1994), contains a spoken language part. Spoken language corpora are much smaller and fewer than corpora of written language. For instance, the British National Corpus contains over 90 million words of written language, compared to "mere" 10 million words of spoken English discourse.

The utility of a corpus is increased when it has been annotated. The information implicitly present in a body of text has thus been made explicit and further enriched. Work on the annotating of spoken language corpora and further application in various practical areas of research gained momentum in the 1980s and especially in the 1990s. In speech corpora, great effort has been expended

on segmental labeling. Gradually researchers have come to realize that much more work is required on the prosodic aspects of speech. However, prosodic annotation is less straightforward and more impressionistic than segmental annotation, as there are far fewer clear acoustic clues to prosodic phenomena. It cannot be done automatically and requires careful listening by a trained ear. It is also necessary to check on the consistency of transcriptions. Spoken language corpora annotated for prosody are thus comparatively small. Degree of phonetic detail offered by particular corpora varies. For instance, the MARSEC corpus works with a transcription system containing far more phonetic data than does the ToBI system (see further). The spoken part of the BNC uses a broad orthographic transcription scheme with little prosodic information.

Reusability of resources and a common interchange format are crucial questions in the research community working on spoken language corpora. In the early 1990s a research project TEI (Text Encoding Initiative) was launched to work towards common guidelines. For computer readability and transfer of both segmental and suprasegmental transcription, various systems have been introduced. In the late 1980s, SAMPA (Speech Assessment Methods Phonetic Alphabet) was created for segmental transcription, which is commonly used in electronicaly published works. A parallel system of prosodic notation, SAMPROSA (SAM Prosodic Alphabet), was later proposed. It was meant for application in multi-tier transcription and representation systems. In the USA, the ToBI (Tones and Break Indices) system was developed (Silverman et al. 1992). The originators of ToBI proposed a simple system that would be easy to learn, leading to good inter-transcriber consistency. This multi-tiered prosodic transcription scheme has largely been used for English, especially American English.

Most computerized spoken corpora have so far been on English. Languages other than English have received less attention. Nevertheless, the standard-setting initiative within the European Union, known as EAGLES (Expert Advisory Group on Language Engineering Standards), covers a range of EU languages in its recommendations. For German, there is the VERBMOBIL project, a large database of German spontaneous speech, covering a variety of different German speaking styles. Part of these data is prosodically labeled. A

Dutch transcription system ´IPO classification of F_0 patterns´ (´t Hart et al. 1990) proved its usefulness for more than one language (next to Dutch and British English, the model has also been applied to German and Russian).

The specifics of tone languages need quite different treatment. In Mainland China, there are several speech databases being developed. However, only one of them has prosodically labeled data. The first version to make a prosodic labeling system for Standard Chinese is the ´National Database Design and Prosodic Labeling for Speech Synthesis´ (later refered to as the NDD). It was developed by the Institute of Linguistics, Chinese Academy of Social Sciences, Peking (Li Aijun, Zu Yiqing) in cooperation with the Dept. of Linguistics, MIT (Li Zhiqiang). Its design was introduced at Oriental COCOSDA (the International Committee for the Co-ordination and Standardisation of Speech Databases) in 1999 (Li 1999). It comprises four sub-databases, one of them being a prosodically labeled dialogue database. This sub-database contains 52 dialogues on ordinary life topics (booking tickets, asking the time, etc.). The multi-tiered prosodic labeling system is based on ToBI, adopting a few symbols from SAMPROSA and adjusting it to the specifics of a tone language. It has not yet assumed its final shape. Another project - ´Speech Database of Chinese Discourse and its Labeling´ - is being currently developed in Peking (ILCASS). It includes syntactic labeling, segmental labeling and prosodic labeling.

2. OBJECTIVES

The main applications for prosodically annotated databases are in two areas, i.e. basic linguistic research on the one hand, and experimental phonetic research and speech technology on the other (the results have then been used in speech synthesis and recognition).

The PALM project (ProsodicAl Labeling for Mandarin) presented here is not meant for speech synthesis. It does not aspire to capture detailed features and tricks of spontaneous speech. It aims to reveal the basic anatomy of the rhythm of spoken Mandarin and its links to the grammar and tempo. One of the means used is statistical processing of the annotated database. Discovered tendencies, patterns and features should find use particularly in the process of teaching Chinese as a second language. The system of prosodic transcription was

thus designed with didactic concerns in mind - the graphic representation is kept as user-friendly as possible. However, its application in teaching practice requires certain adjustments, namely reducing the levels of stress prominency from six to four (see chapter 5.3.3). The scheme can be used either as such for an abstract notation of MC sentences, or converted into a prosodically enriched orthographic transcription based on *pinyin.*

The present paper was compiled by two authors. Hana Třísková designed the system of prosodic transcription and the software and is responsible for the prosodic labeling. David Sehnal did grammatical tagging and used the software for further analysis, aimed at finding correspondence between the syntactic and prosodic levels of Mandarin. The objectives of the project are naturally related to the fact the former author teaches the phonetics of Mandarin to university students.

3. CORPUS

The character of a speech corpus may vary to a great extent: it could be an intimate dialog between friends with typical features of informal speech, such as repetitions, hesitations, incomplete words and self-corrections, overlaps, laughter etc. Or it could be a prepared monolog of a very formal kind. Our database is based on a monological text - a fable story *"Houzi lao yueliang"*- Monkeys Fishing for the Moon (the story comes from the introductory part of the Learning Dictionary of Modern Chinese by O. Švarný, 1998-2000). The text comprises 23 sentences. It was read by a native Pekinese in a rather spontaneous fashion twice: in slow and in fast tempo. (The speech database is thus divided into two parts: the slow one is called ´unit 1´, the fast one is called ´unit 2´.) This allows observing the influence of tempo on the rhythmical scanning of the utterances. As our text is monological read speech with no turn-taking, overlaps etc., it is reasonably straightforward to process. It was recorded in a studio in controlled conditions.

The question of size is the first problem with the term ´corpus´. Nowadays we associate with a corpus the attribute of being large. However, it is hard to say that there is any agreement about appropriate size. It may range up to tens of millions of words (like the above mentioned BNC or COBUILD corpora). Our database is tiny, thus it is better to call it a sample of speech rather

than a corpus. The present task is restricted to testing the prosodic transcription and tuning the functions of the software.

4. METHODS

Any transcript is far from being an objective and exhaustive reflection of speech events - it is always selective and interpretive. Each application tends to bring its own priorities regarding transcription. The degree of interest in detail reflected in various corpora is variable (this variability of detail naturally correlates not only with the purpose of the corpus, but also with the amount of data transcribed). If the prosodically labeled database is used in speech synthesis and recognition, the prosodic system requires a lot of detail. Then it is usually necessary to undertake a full multi-tier prosodic transcription. The transcriber should have access to the F_0 curve and waveforms (and possibly to the intensity curve) for the speech to be transcribed. For instance, the above-mentioned ToBI and NDD systems work with several simultaneous tiers of phonological information.

However, it is possible to mark some prosodic information even in orthographic transcriptions (an example is the Dutch speech styles corpus). The requirement of incorporating all the necessary prosodic data in a single row of transcript, instead of placing it on separate representational tiers, was adopted for PALM. This decision was tempered by the goals mentioned above - the system was designed to help detection of regularities and patterns of interest without the distraction of irrelevant details (furthermore, the transcript needed to be lucid and easily readable to the human eye). The prosodic notation is concentrated on rhythm, namely the stress prominency of individual syllables, and horizontal segmentation (grouping of the syllables). However, the PALM prosodic transcription cannot be considered an orthographic transcription, as the particular words are represented by abstract alphabetic symbols - one letter stands for one syllable, expressing the degree of its stress prominency. (Neverthelss, the system can be converted into a fully orthographic transcription, as indicated above). In addition to the prosodic transcript, the orthographic version of the sentence is also supplied. The third tier - the syllabic pattern - contains data which are implicitly present in the prosodic transcript and has only an auxiliary function. To sum up, although there are three parallel

"tiers", the transcription itself is not actually multi-tiered, as all prosodic information (except for lexical tones, which are found in the orthographic tier) is embedded in a single row of symbols. The system of prosodic transcription and other theoretical concepts draw on the results of previous research conducted by O. Švarný (1974, 1991).

The formulation of the notation system was dictated by the need for learnability and consistency in use, leading to a good inter-transcriber consistency. It needs to be human-readable, as well as machine-readable. Presumably, it should not require great effort to teach it to novice transcribers (if the transcription is used in the teaching process, the users would be newly trained students rather than experts). These assumptions still have to be tested, though. The material was prosodically annotated in an auditory fashion. Various attributes and features were annotated above the transcript itself. These data are presented in a table format (one table for each sentence - cf. Fig.1). First of all, the lexical entries were labeled for the degree of stress prominency of individual syllables, and for the degree of tightness of the bond with neighbouring entries. Furthermore they were labeled for the following characteristics: number of syllables, stress pattern (for compounds), presence of sentence stress, and position within the rhythmical structure of the sentence. Grammatical features (grammatical functional characteristics and syntactical function) were also labeled to allow exploration of links between prosody and grammar.

Let us briefly compare the multi-tier NDD system and PALM system. The former labels the phonetic features with functional significance at five tiers: orthographic tier, tone and intonation tier, sentence function tier, break index tier, and stress/prominency tier. Below follows a comparison of attributes labeled by both systems:

At ´orthographic tier´, both systems give *pinyin* representation including lexical tones (number in the NDD, tone mark in PALM). The NDD also offers Chinese characters and English translation. At ´tone and intonation tier´, the NDD marks tonal features, change of register and range. PALM only marks lexical tones. At ´sentence function tier´, the NDD annotates the following types: S statement, Q interrogative, I imperative, E exclamation. In PALM notation, this attribute is not explicitly annotated. However, sentence function can be distinguished by the sentence punctuation mark (. ? !). At ´break index

tier´, the NDD assumes the following prosodic structure: syllable (S), prosodic word (PW), minor phrase (MIP), major phrase (MAP), intonation utterance (IU). The actually labeled breaks are: minor phrase break, major phrase break, sentence break. PALM works with following hierarchy: syllable, minor tone-unit (MiTU), major tone-unit (MaTU), sentence (for hierarchy of breaks see chapter 5.3.3). At ´stress/prominency tier´, the NDD system marks sentence stress, distinguishing between ´normal´ stress and contrast (emphasis) stress. PALM transcription goes to more detail, marking stress value for each syllable. At the level of words, stress patterns of compounds are annotated. At the sentence level, contrastive/logical stress is marked (see chapter 5.3.3).

The PALM software was developed for the operation system Microsoft Windows 95/98/NT, using the "Delphi 4" programming tool of the Inprise company. The data were saved in Paradox format.

5. DATABASE OF SENTENCES

The database consists of 23 sentences in the slow unit (unit 1) and 23 sentences in the fast unit (unit 2). Altogether there are 358 words (179 in each unit) and 478 syllables (239 in each unit). Each sentence occupies a separate screen, having a unique numeric identifier (cf. Fig.1).

5.1 Units of Analysis

In speech there is basically a continuum without a clear division into units. Thus, one important issue in transcribing speech corpus is segmentation into sentences, intonation units or other units. Researchers use different criteria for dividing text into segments useful for coding and analysis. In our case, as our sample is read speech, the basic units are pre-determined as orthographic sentences[2]. Items subjected to further analysis are units on two levels: words[3], and syllables. In Mandarin, the syllable appears to be an important unit not only at a phonetic level (functioning as a stress-bearing unit) - it functions also as a semantic unit. It seems useful to bring attention down to the syllabic level and mark all syllables for their stress prominency. In the transcript, compounds are thus decomposed into a sequence of letters, each letter representing one syllable.

5.2 Format

There are several options for spatial arrangement of data on the screen. Transcript can be arranged in a traditional book format, where codes are interleaved with the basic level description (´interspersed´ format). When a more detailed treatment is required, parallel representation is more appropriate rather than embedding all information in a single stream - the specifications are placed on separate tiers (SPS, i.e. ´segment-plus-specification´ format). Such representation allows the user to concentrate on specific level/features without distractions. For instance, ToBI uses tonal tier, break index tier and miscellaneous tier. The tiers used in the NDD database were described above. Yet another option is to abandon the familiar horizontal format and present data in a vertical format. This solution was adopted in MARSEC, where data are stored in a relational database and presented in a table format on the screen: *"Book format is an attribute not of speech, but of Western writing systems... since there are often several annotations relating to the same piece of data, book format is in many cases inappropriate."* [4]

The PALM scheme offers both formats. A horizontal format is preserved for the orthographic tier, transcript (prosodic) tier, and syllabic tier. These are placed in the upper portion of the screen (cf. Fig.1).

The orthographic tier ´Sentence´ contains the wording of the sentence in *pinyin* including lexical tones (tone marks are used rather than numbers). The word boundary is indicated by a full stop to allow automatic segmentation. The tier ´Prosody´ contains the prosodic transcript of the sentence. The tier ´Syllables´ contains the syllabic pattern of the sentence (each numeral indicates the number of syllables of a corresponding word; horizontal segmentation is also marked here). This information is of course implicitly contained in the prosodic tier. However, it is marked explicitly here to give a clear cue for retrieval (the number of syllables being one of the retrieval criteria – see chapter 6.1). Furthermore, if the software is developed to analyze higher levels in the future, this cell will also facilitate exploration of the inner structure of minor tone-units. The horizontal tiers are filled in manually. Horizontal format is then converted into a vertical format - the sentence is automatically segmented into words, which are listed vertically in a table as separate entries. These lexical entries are labeled for further features.

Fig. 1   Database of sentences (unit 1, sent. No. 8)

5.3 Prosodic Transcription

The system of prosodic transcription was inspired by O. Švarný´s transcription scheme (Švarný 1991). However, the stress prominency scale is built on somewhat different principles.

5.3.1 Degree of Detail

Transcription is the process of representing spoken language in written (and possibly machine-readable) form. Conversion into a different medium by definition represents a major difficulty. Any specific system of transcription can represent only a select amount of information out of speech signal. Consequently, every transcript is fundamentally selective and interpretive, involving a great deal of subjective choice. As a result, we can observe variable degrees of interest in detail in various transcriptions (as transcribing is extremely time-consuming, there tends naturally to be a certain trade-off between quality, i.e. degree of detail, and quantity, i.e. size of corpus). The design of any transcription scheme should always closely reflect the researcher´s goals. Only then can a transcript help in finding patterns of interest without drowning them in a welter of irrelevant detail - it is not necessarily true that the more elaborate transcription, the better it is. However, one can easily be led into the temptation to believe that the greater the detail, the better the "map". This is neatly illustrated by following quotation from Lewis Carroll´s novel (´Sylvie and Bruno Concluded´, 1893) [5]:

> ´What do you consider the *largest* map that would really be useful?´ ´About six inches to the mile.´ ´Only *six inches*!´ exclaimed Mein Herr. ´We very soon got to six yards to the mile. Then we tried a *hundred* yards to the mile. And then came the grandest idea of all. We actually made a map of the country on the scale of a *mile to mile´*. ´Have you ever used it much?´ I enquired. ´It has never been spread out yet,´ said Mein Herr. ´The farmers objected: they said it would cover the whole country, and shut out the sunlight! So we now use the country itself as its own map, and I assure you it does nearly as well.´

When the transcription scheme strives to capture informal spontaneous speech, beyond the linguistic features attention must be paid to a broad variety of other features (discourse phenomena, the participants´ knowledge and attitudes, paralanguage and kinesic features etc.). In such cases, transcription involves the problems of capturing not only WHO said WHAT to WHOM, but also in what manner, under what circumstances etc. HOW something is said is often much more important than WHAT is said, as all of us know from our own experience. However essential to an understanding such features are, some descriptions here must remain intuitive and speculative. Trying to capture these characteristics is truly an attempt at "transcribing the untranscribable" (to use G. Cook´s formulation – cf. note 1).

Finding solutions to the problem of how to represent non-verbal aspects of discourse and paralinguistic features in visual form is one of the burning issues in handling speech corpora. The importance of paralinguistic and other mentioned features must not be minimized. However, they are less central to the present goals of PALM than the basic distinctions and few major characteristic, which underlie the rhythm of Mandarin speech.

5.3.2 User-friendliness

Transcription conventions always involve using various visual forms (such as capitalization, spacing, hyphens, use of brackets, etc.) for capturing different features of spoken language in a machine-readable form. However, there are different requirements for the transcriptions for human eyes compared to those suitable for machines. Schemes meant for human eyes should be user-friendly. Iconic representations can greatly help. Besides it is wise if the choice of conventions draws upon the reader's expectations and reading habits already formed by literacy.

These concerns were taken into account in designing the PALM notation (e.g. the stressed syllables are always represented by a capital letter, which makes them very easy to spot).

5.3.3 Prosodic Features

Stress patterns and grouping of the syllables within an utterance are the major prosodic features bringing about a sense of rhythm. The resulting

rhythmical form of an utterance is conditioned by a combination of various effects and factors, functioning at different linguistic levels (Třísková 1995).

At the word level, morphological and other properties of words play their role. The grammatical functional characteristic (*yufa gongneng tezheng*) of a particular word is among the relevant factors. For instance, personal pronouns are often unstressed (be it in the function of subject, object or attribute), joining other words as clitics. Grammatical and modal particles invariably behave like enclitics. Full words, on the other hand, are suited to bear stress.

Another factor at the word level is the number of syllables of the word. Monosyllabic words rarely stand alone; they usually join to form larger rhythmical units. Some of them typically receive no stress, behaving as proclitis or enclitics (monosyllabic personal pronouns, prepositions, postpositions, adverbs, grammatical and modal particles etc.). As for compounds, they have no fixed word stress. The exception are disyllabic words with a fixed *qingsheng* on their second syllable (*háizi, gēge*). The rest of the disyllabic words, due to their morphological structure and other factors, have inclination to certain accentuation patterns with a different degree of stability (Švarný 1974). Accentuation of trisyllabic words is less variable.

At the sentence level, the syntactic function of a word within a given sentence plays an important role influencing its prosodic features (e.g. nominal object is typically stressed in neutral speech). Sentence stress is decided in accordance with the semantic structure of the sentence, the communicative intentions of the speaker, his attitudes and moods etc. The position of a word within the sentence plays its role, too, e.g. items placed at the end of intonation units are inclined to bear stress (´end-arsis tendency´ - Švarný 1991).

Rhythmical concerns and changes in tempo can cause redistribution of stress as well as changes in grouping (increased tempo results in a smaller number of rhythmical units of increased length).

Horizontal segmentation

PALM works with a hierarchically organized prosodic structure consisting of non-overlapping units, each of which corresponds to one or more units on the level immediately higher. The adopted hierarchy of the rhythmical units is as follows:

0   syllable
1   minor tone-unit (MiTU)
2   major tone-unit (MaTU)
3   sentence

A <u>minor tone-unit</u> (corresponding to Švarný´s term ´rhythmical segment´ - Švarný 1991) is a sequence of syllables constituting a rhythmical unit with no perceived boundary separating the syllables (the tightness of bond between particular syllables within a MiTU may vary, though). The average number of syllables in a MiTU is 2.5 to 4 syllables, varying between 1 and 10 syllables. Two adjacent MiTUs are separated by a juncture, bearing weak intonational marking. The segmentation of the speech flow into MiTUs is rather unstable and can change with tempo or under the influence of other factors (MiTU boundaries are sometimes hard to determine). In the transcript, syllables constituting a single compound word are represented as an uninterrupted sequence of letters (e.g. Aoa). If a MiTU comprises more than one word, the symbols are double hyphenated (e.g. t--Ao)[6]. Furthermore, a MiTU boundary is signaled by a space (e.g. A--0 t--Ao).

Two or more MiTUs usually join together forming a <u>major tone unit</u> (corresponding to Švarný´s term ´colon´ - Švarný 1991). A MaTU is a rhythmical unit with a distinct intonation, marked by a strong perceptual boundary, with or without a silent pause, lengthening or change in speech tempo. It may correspond to a grammatical sentence, but quite often to a clause or phrase. The average number of syllables is 6 to 7 syllables, varying between 1 and 17 syllables. The segmentation of the speech flow into MaTUs is fairly stable, as it is primarily decided by the grammatical structure of the sentence.

Three types of MaTU boundary are distinguished in PALM transcription:
1.   MaTU boundary signaled by intonational prolongation and perceived disjuncture, but with no terminal intonation and no silent pause, e.g. after a long subject (marked by single comma in transcript).
2.   MaTU boundary signaled by intonational prolongation and a perceived silent pause, but no terminal intonation, e.g. at the end of a non-terminal clause (marked by double comma in transcript).

3. MaTU boundary signaled by terminal intonation and a silent pause, identical to the sentence boundary - full intonation phrase boundary, e.g. at the end of a sentence (marked by sentence punctuation marks . ! ? in transcript)[7].

Silent pauses of course come in a continuum of lengths. This continuum is segmented in various ways by different authors. Pauses can be measured in physical time, or in number of beats relative to speaker's articulation rate. The PALM transcription scheme does not introduce any measurements for silent pauses.

Stress

Stress is graded rather than discrete. To segment and describe this continuum, three binary features were picked up: presence/absence of lexical tone (decided at the word level), presence/absence of ictus[8] (decided at the word level and MiTU level), presence/absence of sentence stress (decided at the sentence level).

1. A straight binary contrast of stressed (i.e. ictus-bearing) syllables vs. unstressed syllables can be considered the fundamental ´philosophy´ underlying the stress scale adopted for PALM. This basic evaluation of stress is decided by comparing the syllable in question to the immediately adjacent syllable/s within the MiTU. The stressed syllable receives MiTU stress. Each MiTU contains one primary stress, quite frequently also one secondary stress[9] (in such case, the most frequent pattern for 3-4 syllabic MiTUs is the ´acronymic´ one, with the stresses falling on first and last syllables - Švarný 1991). Long MiTUs with three stressed syllables are rare, occassionally appearing only in fast speech. If the stressed syllables within a MiTU have a perceivable difference in the degree of their prominency (and the more prominent one is not marked for MaTU stress), primary (A) and secondary (a) MiTU stress can be distinguished in the transcript (e.g. ao--o--A). This yields a finer prominency distinction for stressed syllables.

2. Tonality feature is employed since tone serves as a major vehicle of stress in Mandarin. Stressed syllables as a rule have lexical tone[10]. With regard to unstressed syllables, part of them have underlying tones which become either

15

weakened, or fully neutralized (´tonally neutralized syllables´) in the speech flow. Some of the unstressed syllables, though, have no underlying lexical tone at all, e.g. *le, zhe, guo,* dong*xi* (´toneless syllables´). Their lack of stress is decided at the word level and can be predicted. The PALM transcription distinguishes between tonally neutralized syllables and toneless syllables, offering a finer stress prominency distinction for atonic syllables (the broader term ´atonic syllables´ embraces both ´toneless´ and ´tonally neutralized´ syllables). This distinction is very important, as the neutralized syllables can restore their tone if the prosodic context changes.

     3. One (rarely more) syllable of MaTU can receive an enhanced degree of prominency, bearing either normal sentence stress, or contrastive/logical stress. This kind of stress - <u>sentence stress</u> - is one of the important prosodic features of a minimal utterance. Marking sentence stress yields a finer prominency distinction for stressed syllables.

     The particular levels of stress prominency can be described as follows[11]:

1) Stressed (= ictus-bearing) syllable, enhanced promin. - *sentence stress*   Á
2) Stressed (= ictus-bearing) syllable - *primary MiTU stress*   A
3) Stressed (= ictus-bearing) syllable - *secondary MiTU stress*   a
4) Unstressed syllable with an underlying lexical tone; weak tone   t
5) Unstressed syllable with an underlying lexical tone; neutralized tone   0
6) Unstressed syllable with no underlying lexical tone (i.e. toneless)   o

     The above described three binary features[12] (i.e. presence/absence of lexical tone, ictus, and sentence stress) can be used to code fields with the Boolean values ´True´ (+) or ´False´ (-) and entered in a table:

	Lexical tone	Ictus	Sentence stress
Á	+	+	+
A	+	+	-
a	+	+	-
t	+	-	-
0	+	-	-
o	-	-	-

For didactic purposes, four levels of prominency are sufficient: Á, A, t, o. Such simplified scheme of stress prominency, omitting ´a´ and ´0´ values (merging the ´a´ value into the ´A´ value, and the ´0´ value into the ´t´ value), produces a table with unambiguous distinctions.

The values ´A´, ´t´, ´o´ of this simplified scheme are analogical to three basic types of syllables distinguished by their suprasegmental features as established by Kratochvil (1968) (our ´Á´ value would thus be merged into ´A´ value).

As stress is a relational property, the symbols in the scale express the relative stress prominency of a syllable. The values are context dependent, not absolute, and do not reflect intonational pitch changes. Consequently, the presented scale is not to be considered as a scale of objective acoustic prominency. The elements which happen to be marked identically in the transcript are not necessarily of similar physical parameters and can be perceived as unequally ´strong´, if compared to each other. On the other hand, some of the elements which are marked by different symbols could be perceived as equally ´strong´ in mutual comparison.

In the example below is shown how the above described symbols are used in prosodic transcription.

Example of a prosodic transcript (unit 1, sentence No.14):

Dà.hóuzi,.   yě.dàoguà.zhe,,.   tā.lā.zhe.   lìng.yī.gè.   hóuzi.de.jiǎo.
t--Ao,   t--Át--o,,   t--A--o   Á--t--0   ao--o--A.

Intonation

Intonation units can always be identified not only by syntax or meaning, but also by their auditory properties. Their boundaries are signaled by a combination of prosodic features such as terminal pitch contours, resetting of the pitch baseline, pauses, or changes in tempo. As mentioned above, the present analysis does not employ any explicit ways of rendering the intonation contours.

However, there are implicit options created by use of punctuation marks. Punctuation marks are used in written texts to mark continuing, terminating, questioning, exclamatory intonation etc. Literate people are accustomed to interpret punctuation in that way while reading. A transcription scheme can take advantage of these habits. In the PALM transcription, punctuation marks have a functional role, indicating above described several kinds of MaTU boundary. Of course, the proper interpretation of this implicit marking presupposes a knowledge of the rules governing the intonation of Mandarin, the intonation patterns and their interplay with tones. Punctuation marks are used in the following manner:

single comma:	corresponds to MaTU boundary 1
double comma:	corresponds to MaTU boundary 2
colon:	corresponds to MaTU boundary 2
full stop, exclam. mark:	correspond to MaTU boundary 3
question mark:	corresponds to MaTU boundary 3

5.4 Annotated Attributes

While making a decision as to what should be included in the corpus above the transcribed text itself, it is clear that, in addition to the actual words spoken and some descriptive information, it is highly desirable to also include material which is to some extent analytical as well as descriptive. In our database the words (i.e. entries) are annotated for several grammatical and prosodic attributes. Altogether, seven attributes were chosen to be labeled (cf. Fig.1), which either reflect or influence the prosodic behavior of words and help to reveal its dependency on the grammar of the sentence, tempo, and position of the word within the sentence (the eighth attribute - lexical tone/s, appear on the entry itself). These seven attributes are referred to as *facets*. Data in facets No.1

and No.2 are input automatically by segmentation of the tiers ´syllables´ and ´prosody´. The rest of data has to be filled in manually. All facets can serve as retrieval criteria.

Description of the facets:

´GFCH´ (facet 1) indicates the <u>grammatical functional characteristics</u>[13] the lexical entry belongs to (part-of-speech tagging is the commonest form of corpus annotation). See chapter 8.

´SF´ (facet 2) indicates the <u>syntactic function</u>[14] of the lexical entry within the given sentence. See chapter 8.

Retrieval based on these two facets allows observation of rhythmical behavior of entries of particular GFCHs and/or particular syntactic functions, depending on the prosodic context and on their position. Furthermore, it enables the user to explore variation of rhythmical features of a particular GFCH depending on a specific SF, and vice versa.

´Syllables´ (facet 3) indicates the <u>number of syllables</u> the entry consists of. Explicit labeling of this attribute allows rapid selection of all entries consisting of a specific number of syllables. Retrieval based on facet 3 permits the user to observe rhythmical properties of monosyllabic, disyllabic etc. words depending on their position and grammar.

´Prominency´ (facet 4) indicates the degree of <u>stress prominency</u> of the individual syllables constituting a particular entry. See chapter 5.3.3.

´Pattern´ (facet 5) indicates a <u>stress pattern</u> of a compound. Retrieval based on facet 5 can be combined with the criterion ´number of syllables´. It allows sorting out all entries sharing the same realization pattern. Most interesting is, first of all, to observe to what extent the accentuation of a particular word is stable in changing contexts. Another option is to explore what words tend to adopt particular rhythmical patterns (depending on their position, GFCH and/or SF).

A large corpus containing numerous occurrences of the one lexical unit would allow us to attempt to categorize compound words according to their accentuation tendencies. Much research has been done on accentuation of compounds by Švarný (for disyllabic compounds he singled out seven categories, termed ´accentuation types´ - Švarný 1974).

The basic stress patterns marked in this facet are four: Mo = monotonic (first syllable tonic, remaining syllable/s atonic, e.g. _**hóu**zi_), Desc = descending (sequence of tonic syllables with descending prominency, e.g. _**dào**guà_), Asc = ascending (sequence of tonic syllables with ascending prominency, e.g. _yì**zhí**_), Akro = acronymic[15] (3-4 syllabic sequences: the first and last syllables are stressed, e.g. _**lǎo**bu**zhǎo**_).

´S. stress´ (facet 6) contains data about <u>sentence stress</u>: the ´+´ symbol is filled in if any of the syllables of the entry in question is marked for sentence stress. Two types of sentence stress are distinguished:

a) Neutral, ´normal´ sentence stress, i.e. the stress assigned by grammar. No word is singled out for particular emphasis. Contrast is not involved. In this case, the sentence stress falls on the last (i.e. the rightmost) word of the sentence. The exception are items which cannot bear normal stress - monosyllabic postpositions (except for _lǐ_), monosyllabic personal pronouns functioning as object, modal particles, grammatical markers etc. When the sentence receives ´normal stress´, it is not marked in PALM, since it is predictable and usually not of especial acoustic prominency.

b) Non-neutral sentence stress, i.e. emphatic, logical or contrastive stress. It can fall on any word (even preposition or conjunction etc.). Non-neutral sentence stress is always marked in PALM.

´Position´ (facet 7) indicates the <u>position of the entry</u> within the rhythmical structure of the whole sentence. This label consists of two pieces of data:

1. The position of the pertinent entry within its ´mother´ minor tone-unit. It is indicated by lower-case letter: a = initial position, b = medial position, c = terminal position, m = MiTU consisting of a single word.

2. The position of the pertinent minor tone-unit within its ´mother´ major tone-unit. It is indicated by upper-case letter: A = initial position, B = medial position, C = terminal position, M = MaTU consisting of a single MiTU)[16]. Retrieval based on facet 7 allows the user to observe the prosodic properties of words sharing the same type of position, or to observe the changes of prosodic features of a particular word depending on its position.

Example of position labels (unit 1, sentence No.8):

	Lǎo	hóuzi	pǎoguòlai	le,,	hòubian	gēn	zhe	yī	qún	hóuzi.
prosod. pattern:	t--	Ao	At0--	o,,	Áo	A--	o--	0--	a	Ao.
position tags:	Aa--Ac	Ca--		Cc,,	Am	Ba--	Bb--Bb--Bc			Cm.

With regard to horizontal segmentation, it is marked not only in the transcript itself - it is also joined to the labels ´syllables´, ´prominency´, and ´position´ in the table (similarly to MARSEC). Hyphenation, signaling the links of the word to adjacent words within a MiTU, is always joined to both sides of the label ´syllables´ and ´prominency´ (e.g. -2- -to-). Sentence punctuation mark is always associated with the label of the preceding word (e.g. *laochulai*! 3! Á00! Mo!).

6. QUERY

The Query screen (cf. Figs. 2, 3) allows the user to ask questions about the data stored in the database. The corpus can be accessed by particular sentence in the Database of sentences, then by particular word, or by a number of features of either words or syllables (Query). Subsets of entries sharing a specific label can be retrieved and displayed.

Furthermore, since retrieval criteria can be combined for words (a set of 2 to 5 criteria), it is possible to retrieve co-occurrences of various prosodic and grammatical features, for instance to find stressed monosyllabic personal pronouns functioning as object, or unstressed disyllabic nouns of Mo type functioning as subject, etc. Since the punctuation and position of the words are also linked, it enables the user to find, for instance, words which are accented before a full stop.

Fig. 2   Example of a query: disyllabic nominals functioning as subject

Unit	Sent.No	Entry	GFCH	SF	Numb. of Syllables	Prominency	Pattern	S.stress	Position
1	02	lǐtou,	Nl	zhu	2,	Ao,	Mo		Crn,
1	04	Yuèliang	N	zhu	2	Áo	Mo	+	Arn
1	07	Yuèliang	N	zhu	2	Áo	Mo	+	Arn
1	08	hòubian	Nl	zhu	2	Áo	Mo	+	Arn
1	09	Tāmen	Np	zhu	2	Ao	Mo		Arn
1	10	Yuèliang	N	zhu	2	Ao	Mo		Aa
1	18	Tāmen	Np	zhu	2	Ao	Mo		Arn
1	21	yuèliang	N	zhu	2	Ao	Mo		Brn
1	23	Yuèliang	N	zhu	2	Ao	Mo		Arn
2	02	lǐtou	Nl	zhu	2-	Ao-	Mo		Ca
2	04	Yuèliang	N	zhu	2	Ao	Mo		Arn
2	07	Yuèliang	N	zhu	2	Ao	Mo		Arn
2	08	hòubian	Nl	zhu	2	Ao	Mo		Arn
2	09	Tāmen	Np	zhu	2-	ao-	Mo		Ma
2	10	Yuèliang	N	zhu	2	Ao	Mo		Arn
2	18	Tāmen	Np	zhu	2	Ao	Mo		Arn
2	21	yuèliang	N	zhu	-2	-Ao	Mo		Ac
2	23	Yuèliang	N	zhu	2	Ao	Mo		Arn

6.1 Queries About Words

Queries about words can search the database in two ways:

a) All occurrences of a particular word in the database can be compared by entering the word in the cell ´entry´ (i.e., the user can check how the chosen word behaves in different contexts). This can be combined with the retrieval criteria ´syntactic function´, ´pattern´, ´position´, and ´unit´. This function allows observing the changes of rhythmical behavior of the word (as a type), depending on its grammatical and syntactic characteristics, position within the sentence, and prosodic context.

b) The subsets of entries meeting particular retrieval criteria can be sorted out (cf. Fig.2.: disyllabic nominals functioning as subject). Retrieval criteria correspond to the facets from the database. They can be entered in cells A, B, C, D, E:

(A) Number of syllables of the entry (Field list. Options: 1, 2, 3, 4)

(B) Grammatical functional characteristics (Options are listed in chapter 8)

(C) Syntactic function (Options are listed in chapter 8)

(D) Pattern (Field list. Options: Mo, Asc, Desc, Akro)

(E) Position. (See chapter 5.4) Broader retrieval conditions can be defined by replacing any of the symbols by ´%´ symbol (e.g. C%!, instead of Cm!). Thus, for instance, all items followed by a particular punctuation mark can be retrieved (e.g. %%!)

If the user wants to explore the stress features of monosyllabic words (i.e. (A) = 1), then above described criteria - except for (D), of course - can be further combined either with criterion (F), or (G) (see further).

6.2 Queries About Syllables

Queries about syllables allow observing the tonality features, stress features and the relationship between tone and stress (cf. Fig.3: unstressed monosyllabic entries from unit 1). Criteria concerning syllables can be entered in the field lists either F, or G (properties of monosyllabic words can be explored by combining it with the criterion ´number of syllables´ (A) = 1).

Fig. 3   Example of a query: unstressed monosyllabic words from unit 1

(F) Prominency. Retrieving in this cell allows finding all syllables sharing the same value of stress prominency. Any of the six values of the prominency scale can be entered here (Á, A, a, t, 0, o).

(G) Features. The six values offered by this field list allow retrieving the syllables sharing one of the following six characteristics: arses, theses, tonic, atonic, proclitics, enclitics. These options express following features:

- Presence/absence of ictus (i.e. the features ´arses´= stressed syllables, ´theses´ = unstressed syllables) E.g. the ratio of stressed vs. unstressed monosyllabic words can be compared. Compound entries containing such syllable can be searched, too.

- Presence/absence of tone (i.e. the features ´tonic´, ´atonic´). Compound entries containing such syllable can be searched, too. Note that ´atonic syllables´ include both tonally neutralized syllables (0) and toneless syllables (o) (for finer distinction between 0, o the cell ´prominency´ (F) has to be used instead).

- Initial/terminal position of the unstressed monosyllabic entry within the MiTU (i.e. the features ´proclitics´ = initial position, ´enclitics´ = terminal position). This feature allows finer retrieval of the unstressed monosyllabic entries based on their position. E.g., by means of double-step retrieval, unstressed clitical entries can be contrasted with remaining unstressed monosyllabic entries (i.e. those embedded within the MiTU). Another option is to sort out enclitical syllables bearing a weak tone in contrast to toneless enclitical syllables, etc.

6.3 Size of the Subset

The size of the resulting subset always appears automatically in the upper middle part of the screen. It is given both in absolute number and in percentage. Count function (´entries total´ and ´syllables total´) tallies the number of entries in an underlying query. The percentage function expresses the ratio of the number of entries in the subset to the total number of entries in the respective unit. Statistical processing of the labeled entries provides a basis for further analysis, the aim of which is to discover general tendencies.

6.4 Number of the Unit

As mentioned above, the database contains two parallel units: the text recorded in a slow tempo (unit 1) and the same text recorded in a faster tempo

(unit 2). If the number of particular unit is entered, comparison of both units can be made in order to observe changes of various phenomena in dependence on tempo.

6.5 Result Set

The subset of entries corresponding to the results of particular query appears in a row-and-column format in the lower portion of the screen (the arrangement of the table is analogical to the tables in the ´Database of sentences´).

Once the subset is displayed, double clicking on a particular row in a table displays the underlying sentence in the ´Database of sentences´ where the entry comes from. This allows the user to see the whole sentence context.

7. EXAMPLES OF ANALYSIS

7.1 Tone, Stress and Tempo

Results of three queries about syllables are given below to show how the PALM software works. All three queries are focused on comparison of slow and fast speech tempo. Since the quoted numbers are based on a small sample, they do not admit interpretation - the aim is to show the possibilities of the software. Both examples of analysis (7.1, 7.2) and the comments are meant as suggestions for future research based on a large annotated corpus.

The contrast of stressed and unstressed syllables is the major feature underlying the PALM prosodic transcription. Stress characteristics is examined in table No.1.:

Table No.1 – Ratio of stressed and unstressed syllables in the speech flow

	UNIT 1 (slow)	UNIT 2 (fast)
arses (stressed syllables)	42 %	41 %
theses (unstressed syllables)	58 %	59 %
syllables total	*100 %*	*100 %*

With increased speech tempo, the number of stressed syllables decreases slightly. However, the decrease is almost negligible in case of our text. This leads to the assumption the ratio of stressed to unstressed syllables might not be considerably dependent on speech tempo.

Presence/absence of tone is a feature which cuts partly across the stress feature: while stressed syllables always bear tone (with above mentioned exceptions), unstressed syllables sometimes do, sometimes do not. Tonal characteristics is examined in the following table:

Table No.2 – Ratio of tonic and atonic syllables in the speech flow

	UNIT 1 (slow)	UNIT 2 (fast)
tonic syllables	62 %	57 %
atonic syllables	38 %	43 %
syllables total	*100 %*	*100 %*

The ratio of atonic syllables grows with increased tempo. It follows that, in faster speech, some syllables with underlying tones lose their tone and become neutralized (the ratio of toneless syllables remaining constant by definition).

If we combine both features (stress and tone), we arrive at a finer scale of stress prominency for unstressed syllables. As described above, three levels of stress prominency for unstressed syllables are distinguished: tonic (t), neutralized (0), toneless (o). Their ratio is compared in table No.3:

Table No.3 – Ratio of tonic and atonic syllables within unstressed syllables

THESES (unstressed syllables)		UNIT 1 (slow)	UNIT 2 (fast)
tonic theses (t)		20 %	17 %
atonic theses	neutralized (0)	11 %	15 %
	toneless (o)	27 %	27 %
theses total		*58 %*	*59 %*

We can observe, that toneless syllables clearly dominate (their percentage remaining constant by definition) over the remaining two groups of

unstressed syllables. The percentage of tonally neutralized syllables remains the smallest in both units. The difference between the number of neutralized theses and tonic theses becomes less dramatic as speech tempo increases, since some of originally tonic theses lose their tone.

7.2 Relationship between Prosody and Grammar

It is commonly acknowledged that syntactic structure of a sentence plays an important role in constituting the final prosodic shape of the corresponding utterance. The PALM software enables to find evidence of the relationship between prosody and grammar. In our database we used the method of IC analysis of the Peking University linguistic school as developed by Prof. Zhu Dexi and Prof. Lu Jianming. Each entry in the database was labeled for part of sentence characteristic (*jufa chengfen*). This characteristic is, in fact, syntactic function of a particular minimal syntactic unit which enters an IC pair on the lowest level of syntactic analysis. The abbreviations listed in the followig table apear in the facet 2 (synatctic function, SF).

List of syntactic functions:

Syntactic function	*Abbreviation*
Subject	*zhu*
Predicate	*wei*
Predicator	*shu*
Object	*bin*
Complement	*bu*
Attribute	*ding*
Adverbial modifier	*zhua*
Head	*zhong*
Conexial element	*jian*
Enclitical Auxiliary Word	*jia*
Element to which *jia* is attached	*fu*
Construction with Coupled Verbs	*lian-dong*

Furthermore, part of speech characteristic (*cilei*) was labeled, i.e. grammatical functional characteristic *(yufa gongneng tezheng)* which the word represented by a particular entry belongs to. Part of speech is understood as a sum of possible relevant syntactic functions which a word may carry out in the whole grammatical system. In principle, we adopted the Peking University scheme here as well. The abbreviations listed in the followig table apear in the facet 1 (GFCH).

Lists of grammatical fuctional characteristics (bold ones occur in our sample):

GFCH	Abbreviation
Noun	N
Locative Noun	Nl
Temporal Noun	Nt
Pronoun	Np
Deictic Word	D
Measure Unit	U
Classifier	Uc
Nominal Measure Unit	Un
Numeral	C
Verb	V
Intransitive Verb	Vi
Transitive Verb	Vt
Verb of Existence	Ve
Verb of Identification	Va
Prepositional Verb	Vp
Postpositional Verb	Vpv
Modal or Phasal Verb	Vv
Adjective	Adj
Pro-predicative (daiweici)	Dwc
Adverb	Adv
Attributive Word	Atr
Stative Word	S

Postposition	Pp
Conjunction	Co
Verbal Marker	Mv
Other Marker	M
Modal Particle	P
Exclamatory Word	Ex
Onomatopoeia	On

Syntactic functions occur within the scope of IC analysis always in pairs, forming a syntactic structure, or syntactic construction. The order and the set of the functions in particular type of construction is as a rule constant. Below are listed the relevant binary constructions, as they appear in our example sentence (Fig. 4):

List of relevant syntactic constructions:

递进复句	Progressive Complex Sentence
并列结构	Coordinated Construction
主谓结构	Subject-Predicate Construction
述宾结构	Verb-Object Construction
介宾结构	Prepositional Construction
定中结构	Attribute-Head Construction
状中结构	Adverbial Modifier - Head Construction
( )	Marker of Enclitical Auxiliary Element

The adopted system of the IC analysis which constitutes the basis for further discussion is illustrated by the following example - Fig.4 (the characters highlighted by black boxes represent syllables bearing arsis and the corresponding syntactic functions of entries in which these syllables occur).
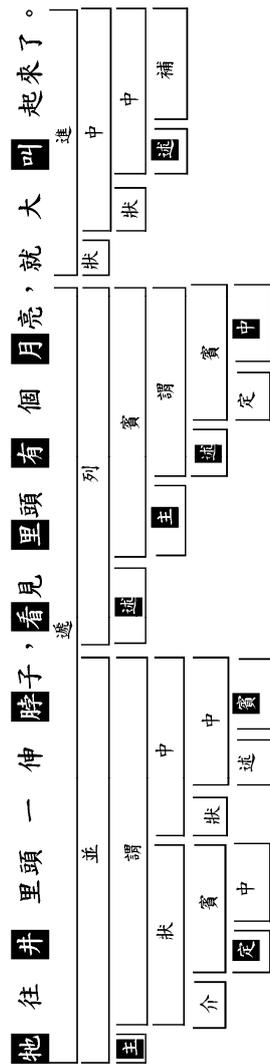
他住井里頭一伸脖子，看見里頭有個月亮，就大叫起來了。

Fig. 4 Example of syntactic analysis (sentence No.2)

Statistic results

Below are given examples of possible application of the PALM software for analyzing the links between prosodic and grammatical features.

For the following statistics we have chosen the contrast between entries which contain at least one stressed syllable (i.e. one arsis), and entries which do not contain any stressed syllable (regardless of number of syllables the entry consists of).

For searching were used the values ´arsis´ and ´thesis´ from criterion (G) = ´features´. Subsequently the correspondence between A+ / A- value of the entries and their grammatical functions were examined.

A+   entry containing at least one arsis
A-   entry containing no arsis

The following table shows the ratio of A+ entries and A- entries in unit 1 and unit 2:

Table No.4 - Ratio of A+ entries and A- entries for both units

	UNIT 1 (slow)	UNIT 2 (fast)
A+ entries	54 % (96 entries)	52 % (93 entries)
A- entries	46 % (83 entries)	48 % (86 entries)
entries total	*100 % (179 entries)*	*100 % (179 entries)*

Obviously, there is no significant difference between the ratio of A+ entries in unit 1 and in unit 2. This result corresponds with almost negligible increase of theses in unit 2 as shown in Table No.1.

Since the difference between the two units regarding the amount of A+ entries is not significant, the next table deals only with unit 2 which was read in more natural manner. It contains information about the distribution of A+ entries among the most important GFCHs, as they occur in our sample:

Table No.5 - Ratio of A+ entries for particular GFCHs in unit 2

GFCH	GFCH total	GFCH A+	Rate of GFCH A+ to GFCH total	Rate of GFCH A+ to A+ total
N + Nl	51	39	76 %	42 %
V	50	36	72 %	40 %
Adj	18	6	33 %	6 %
Adv	16	5	31 %	5 %
U	8	3	38 %	3 %
Atr	1	1	100 %	1 %
D	2	1	50 %	1 %
C	4	1	25 %	1 %
Dwc	1	1	100 %	1 %

The majority of words in our text are nominals and verbs (see the first two rows of the table). They are well suited to bear stress – 76 % of nominals and 72 % of verbs are of A+ type. Nominals make up 42 % of all A+ words, while verbs make up 40 % of all A+ words. It follows that nouns and verbs together cover 82 % of all A+ words in our database.

Out of the adjectives occurring in the database, those functioning as attributes are always of A- type (as was found through means of a combined query). It follows that all A+ adjectives (33%) are used in other functions then attributes (as was found through means of a query combining GFCH and SF retrieval criteria). This result might lead us to presume adjectival attributes are mostly of A- type (in neutral speech). A+ adjectives make up mere 6% of all A+ words.

Adverbs function only as adverbial modifiers in our system. In our sample they are usually of A- type (69 %). They make up only 5 % of all A+ words.

Prosodical labeling of lexical entries and combination of this kind of labeling with the syntactic and lexical features of the entries will permit to postulate more precisely the underlying principles of the IC structure of modern Mandarin. On the level of GFCH, possible regularities or irregularities of stress patterns observed for particular compound words can serve as an important

criterion for distinguishing between disyllabic words and collocations of two monosyllabic words. On the level of SF, postulating such regular patterns could serve as a powerful didactic tool allowing the students to master the prosodic shape of Modern Mandarin, and enabling to study and teach prosody and syntax as language phenomena functioning as an inseparable complex.

9. DISCUSSION

- The system of prosodic transcription has to be checked for inter-transcriber consistency. In the present shape of PALM the evaluations are based solely on the perception - so far instrumental measurements have not been included. Adding at least a fundamental frequency track would offer considerable support for the evaluations, thus greatly improving the quality of labeling.

- Nowadays, digitization technology allows the sound, as well as the transcription, to be computerized. If the user has access to the original audio-recording, then information that may have been lost between recording and transcription can be recovered. For instance, the MARSEC project is designed to align the text of the corpus with the acoustic signal, allowing location of the acoustic information corresponding to a specific piece of the corpus down to the size of individual syllables. We also believe strongly in the value of presenting the acoustic as well as the textual data, permitting the user to listen and look simultaneously. This requirement poses a task for the future development of PALM.

- The presented speech material is a read speech text of a very small size. To yield worth-while results, large speech corpus comprising both read and spontaneous speech needs to be processed.

- For the time being, PALM works on the level of syllables and words. However, the software is open to further development for analysis of higher linguistic levels. Further extension of the functions could mean retrieving from the corpus and concordancing all occurrences of particular strings of elements (such as strings of GFCHs, of SFs, of symbols within the tiers ´prosody´ and ´syllables´). This would allow the user, for instance, to observe the inner structures of MiTUs and MaTUs and to reveal their links to grammar (many observations here were made by O. Švarný (1991, 1998). Significant patterns of

co-occurrence could be discovered. Furthermore sentence functions, sentence intonations and other features on the sentence level could be explicitly labeled.

NOTES

1. Cook, G. 1995. Theoretical issues - transcribing the untranscribable. In Leech, G., Myers, G., and Thomas, J., eds., *Spoken English on Computer*, New York: Longman Publishing, 35-53.

2. The category ´sentence´ in our database is understood as a ´minimal utterance´ characterized by formally and/or prosodically expressed modality, rather then as an independent syntactic unit. In transcript its boundary is signaled by a full stop, exclamation mark, or question mark.

3. The category ´word´ in our database is understood as a syntactic unit, which can constitute a minimal clause and cannot be divided into smaller syntactic elements. For the purposes of our analysis, verbs with modificators (such as *kanjian*) are treated as one word, while aspect markers *le, zhe, guo* are treated separately.

4. Knowles, G. 1995. Converting a corpus into a relational database - SEC becomes MARSEC. In *Spoken English on Computer*, 208-219.

5. Quoted after Cook, G. 1995. Theoretical issues - transcribing the untranscribable. In *Spoken English on Computer*, 35-53.

6. Double hyphen enables preservation of one hyphen with each of the hyphenated elements in the table after the automatic segmentation is carried out.

7. The notation of breaks used in the tier ´syllables´ is analogous to that in the tier ´prosody´.

8. The term ´ictus´ refers to all stressed syllables of the speech flow (regardless of their particular level of stress prominency).

9. While unstressed syllables can neighbour within one MiTU, stressed syllables are not supposed to occur next to each other. However, occasionally two adjacent syllables within one MiTU bear a level of stress which is perceived as equal. A solution for these cases has yet to be found.

10. Occassional exceptions to this rule are some toneless syllables which can become quite prominent under the influence of certain modality (e.g. modal particles).

11. The use of letters ´A´, ´a´, ´t´ is based on Švarný´s terms *´arsis´* for stressed syllables, *´thesis´* for unstressed syllables.

12. The MARSEC transcription scheme describes stress prominency values (´tonetic stress marks´) in terms of six binary features. However, it is applied on stressed syllables only (Knowles, G. 1995. Converting a corpus into a relational database - SEC becomes MARSEC. In *Spoken English on Computer*, 208-219).

13. The GFCH in Mandarin can be defined as sum of syntactic functions the word (lexeme) is able to carry out. It roughly corresponds to the term ´part of speech´ in Indoeuropean languages, however, it is defined exclusively by syntactic means. In our database we basically follow the system of GFCHs established and developed by Prof. Zhu Dexi and Prof. Lu Jianming from Peking University. If necessary, subcategories of GFCHs are marked, e.g. Vt (transitive verb), Nl (place noun) etc.

14. Syntactic functions are realized in pairs of immediate constituents. The IC analysis model currently used by the Peking school of Chinese linguistics, established by Prof. Zhu Dexi, was adopted in the database. Here the ´syntactic function´ of a word means the function which labels the lowest level of syntactic IC analysis of the sentence construction. Higher levels are taken into account only if the word functions as a ´head´ (*zhongxinyu*). In these situations we also mark the syntactic function of the construction the particular word belongs to.

15. Term used by O. Švarný (1991), cf. chapter 5.3.3.

16. E.g. Ca = first minor tone-unit (a) of the terminal major tone-unit (C), Bm = medial minor tone-unit within its ´mother´ major tone-unit (B), consisting of single word (m).

REFERENCES

GIBBON, Daffyd, and RICHTER, Helmut, eds. 1984. *Intonation, accent and rhythm. Studies in discourse phonology*. Berlin: Walter de Gruyter.

GOLDSMITH, John A. 1995. *The handbook of phonological theory*. Blackwell.

HANDEL, Stephen. 1993. *Listening*. Cambridge Mass.: MIT Press.

HOA, Monique. *L´Accentuation en Pekinois*. 1983. Paris: L´Association de Recherches Interculturelles sur l´Expression Gestuelle et Orale.

KRATOCHVIL, Paul. 1968. *The Chinese language today.* London: Hutchinson University Library.

LEECH, Geoffrey, MYERS, Greg, and THOMAS, Jenny, eds. 1995. *Spoken English on Computer: transcription, mark-up and application.* New York: Longman Publishing.

LI, Aijun, ZU, Yiqing, and LI, Zhiqiang. 1999. A national database design and prosodic labeling for speech synthesis. Proceedings of Oriental COCOSDA ´99, Taipei, Taiwan.

SHEN, Xiao-nan Susan. 1990. *The Prosody of Mandarin Chinese.* Berkeley: University of California Press.

SPESHNEV, Nikolai Alekseevich. 1980. *Fonetika kitajskogo jazyka* (Phonetics of Chinese language). Leningrad: Leningrad University Press.

ŠVARNÝ, Oldřich. 1974. Variability of tone prominence in Chinese (Pekinese). In *Asian and African languages in social context*, ed. by L. Hřebíček, 127-186. Praha: Academia.

_____. 1991. The functioning of the prosodic features in Chinese (Pekinese). *Archív Orientální* 2, 208-216.

_____. 1998-2000. *Učební slovník jazyka čínského* (A Learning dictionary of Modern Chinese I-IV). Olomouc: Publishing House of the Palacký University.

ŠVARNÝ, Oldřich, and RUSKOVÁ, Tang Yunling. 1991. Prosodic features in Chinese (Pekinese). Prosodic transcription and statistical tables. *Archív Orientální* 3, 234-254.

TŘÍSKOVÁ, Hana. 1995. Fyziologie rytmu, rytmus mluvené čínštiny (Physiology of rhythm, rhythm of spoken Chinese). Czech version of the paper presented at the 10th EACS conference 1994, Prague. In *O čínské filozofii, literatuře a umění*, ed. by V. Liščák. Praha: Jůza +Jůzová, 51-73.

_____. 1997. Rhythmical behavior of chosen monosyllabic lexical units of spoken Standard Chinese. Paper presented at the 6th ICCL conference, Leiden. Mimeo.

_____. 2000. Degrees of stress prominency in Mandarin – language teaching perspective. Paper presented at the 13th EACS conference, Torino (to appear in the proceedings from the conference).

TSENG, Chiu-yu. 1990. *An Acoustic phonetic study on tones in Mandarin Chinese*. Taipei: Academia Sinica, Institute of History and Philology.

WU, Jiemin. 1998. *Xinbian putonghua jiaocheng*. (A new teaching course of putonghua). Hangzhou: Zhejiang daxue chubanshe.

ZHU, Dexi. 1962. Jufa Jiegou. *Xiandai Hanyu Yufa Yanjiu*. Beijing: Shangwu Yinshuguan, 1985, 42-66.

_____. 1985. *Yufa Dawen*. Beijing: Shangwu Yinshuguan.

_____. 1982. *Yufa Jiangyi*. Beijing: Shangwu Yinshuguan.