ROCZNIK ORIENTALISTYCZNY, T. LXII, Z. 1, 2009, (s. 80–91)

**KRZYSZTOF DYCZKOWSKI, NORBERT KORDEK, PAWEŁ NOWAKOWSKI, KRZYSZTOF STROŃSKI**
(Poznań)

# The Phonetic Grammar of Mandarin Chinese – a Computational Comparative Analysis*

## Abstract

The aim of the paper is to present an interdisciplinary project involving comparative analysis of phonetic systems of Hindi, Mandarin Chinese and Polish. The analysis is conducted with the use of computational tools designed purposely for the needs of phonetic grammar. The basic potential of the computer application and the algorithms employed in it are briefly discussed. The application is intended to enable a uniform description and comparative analysis of many languages. In the initial stage the languages considered are Polish, Chinese and Hindi. The focus of this paper is on Mandarin Chinese. The basic concept of phonetic grammar is introduced, and some specific problems which Chinese presents to the theoretical framework are addressed.

## 1. Phonetic grammar

The idea of using mathematical, logical and computational apparatus to analyze natural language is, of course, neither new nor revealing. Even in the field of phonetic and phonological research alone, there have been a considerable number of proposals in a similar vein to that presented here. In this context we should mention the research of T. Batóg (1967), M. Steffen-Batóg and T. Batóg (1980), W. Meyer-Eppler (1969), J.E. Grimes and E.B. Agard (1959), to name only a few. The limitations of this

* The project is realized by Krzysztof Dyczkowski, Norbert Kordek, Paweł Nowakowski, and Krzysztof Stroński, and is supported by the Ministry of Science and Higher Education grant N N104 327434.

paper do not allow us to elaborate more on those proposals and the differences between those and our framework, so we restrict ourselves to presenting the project in question.

The notion of grammar may be understood in many ways. Probably in the most abstract sense it can be conceived as a calculus of a particular subsystem of a language. In the traditional sense of this term, "phonetic grammar" may sound odd, as "grammar" in this sense is associated with such subsystems as morphology, semantics and syntax, more loosely with phonology and hardly ever with phonetics. This may be a result of the anatomical, acoustical or auditory character of basic phonetic studies, which indeed are not grammatical in nature.

We follow the theoretical proposals of J. Bańczerowski (1985, 1987, 1990, Bańczerowski et. al. 1982) where the results of the basic phonetic research are used as data in a framework that has all the traits of grammatical theory, i.e. it operates on mutually related phonetic units. It is of secondary importance at this point that Bańczerowski's concept of phonetic theory is axiomatic in method. The most relevant property of this proposal is that dependencies and relations between phonetic units are described in the same way, as far as methodology is concerned, as those between morphemes in morphological theory, between words in syntactic theories, etc.

The basics of the theoretical framework which we are adapting to the project are extensively introduced in Bańczerowski 1987 (and more concisely in Bańczerowski et al. 1982). Here we will restrict ourselves to introducing only those theoretical aspects of phonetic grammar which are necessary in order to give the reader a clear understanding of the project.

The articulatory phonetics of a given language is most often understood as a set of phones,[1] each having an articulatory description of some sort. Phonetic grammar is no different in this respect, but extends beyond the mere inventory of phones, treating them as input data for analytical operations which will be briefly introduced in this paper.

The relevance of the phonetic inventory is a property shared by all phonetic studies. Our approach in this respect is unique in taking into account only an anatomical description of the process of articulation. It is probably impossible to completely eliminate the influence of phonological properties of a given language on the shape of the articulatorily driven phone inventory, but nevertheless it is one of the theoretical presumptions of phonetic grammar that it should be kept separate from phonological issues. That assumption presents a problem when it comes to making use of existing inventories of phones of particular languages, Mandarin Chinese being no exception. It is a common practice that elaborating a sound/phone inventory involves minimizing the number of elements in the set. Minimizing usually means taking into account the phonological properties of the system.[2] This kind of procedure is directly opposite to what we assume in our study. The number of elements in the inventory is not important – priority is given to providing a detailed

[1] For the purpose of this survey the term 'phone' may be considered equal to 'sound'. In fact between these two units there exists a type-token kind of difference.

[2] For example, Duanmu 2000.

list of all articulatorily distinctive phones that are found in a particular language. That, for example, includes all distributional variants of what is seen as a single element of the inventory from the phonological point of view. We shall come back to the problem of building a phone inventory for Mandarin Chinese in the last section of the paper.

The process of building an inventory within the framework of phonetic grammar requires assignment of articulatory features to every phone. The operation of assigning articulatory features to a phone is called articulemization. An articulatory feature is a single characteristic of phones in some articulatory aspect. The set of all features is categorized into subsets by the relation of articulatory homogeneity, which holds between features that are characteristics in the same articulatory aspects, in other words features that can be compared. Such subsets are called articulatory dimensions – a dimension is a set of homogeneous features. Below is a preliminary list of dimensions that is sufficient for an exhaustive description of phone inventories of Chinese, Hindi and Polish:

– the mechanism of the air flow origin
– the direction of the air flow
– the state of glottis
– the route of the air flow
– the place of articulation
– the articulator
– **the position of the middle of the tongue**
– the degree of supraglottal aperture
– the vertical position of the tongue
– the horizontal position of the tongue
– the degree of labialization
– the degree of delabialization
– the duration of articulation
– *the degree of supra- and subglottal tension*
– **the slide movement**
– the frequency of articulatory approximation
– **the degree of glottal aperture**
– **the strength of approximation**.

Bold type denotes the dimensions that were added to the original proposal in Bańczerowski 1982, while cursive type denotes a dimension that can be found in the original proposal, but is irrelevant to all languages in the present survey. There are two important properties of articulemization to be mentioned at this point:

– since there is a 'none' feature in each dimension, every phone is assigned a feature in each dimension, even if it is articulatorily unspecified in a given respect[3]
– not only relevant features[4] are assigned to each phone; the number of features of each phone in all languages is the same and is equal to the total number of dimensions.

[3] For example, all vowels in the dimension 'place of articulation'.

[4] The unique set of features that distinguish a given phone from any other.

Another important change to the original framework is the introduction of numerical interpretation of features in each dimension.

D denotes the set of all articulatory dimensions. Each feature is uniquely specified in each dimension by one numerical value from the interval [0,*k*], where *k* is the maximum number of features in a dimension. In that way each phone *f* is specified by a vector, i.e. $f = (c_1,c_2,\ldots,c_n)$ where $c_i$ belongs to the set of articulatory features of a given dimension $D_i$. As an example, in the example dimension below the phone [p] is assigned the value '1', which stands for upperlabiality. The whole vector is coded by repeating the procedure in each dimension. As a result each phone becomes an object in n-dimensional space[5] represented by a vector. Due to limitations of space, we will not give an extensive description of each dimension – we will restrict ourselves to the example of just one of them:

$D_5$: Place of articulation

| n | feature |
|---|---|
| 0 | None |
| 1 | Upperlabiality |
| 2 | Upperdentality |
| 3 | Postdentality |
| 4 | Upperalveolarity |
| 5 | Postalveolarity |
| 6 | Praepalatality |
| 7 | Postpalatality |
| 8 | Velarity |

The numerical values of features in every dimension are not random – for instance in the example dimension the increasing numbers indicate the actual order of the relevant anatomic parts of the oral cavity, from the furthest front to the furthest back.

The phonetic grammar may now be introduced as a set of relations between phones (objects in n-dimensional space) and articulatory dimensions. B a ń c z e r o w s k i's original framework introduced the notion of articulatory distance equivalent to the Hamming distance. This is measured by the number of differential features (features by which one phone differs from the other). For example the Hamming distance between the phones [p] and [b] equals 1, because these phones are differentiated by one feature (or

[5] Where 'n' is simply the number of articulatory dimensions, which is 18 at the present stage.

differ in one dimension). The vector interpretation that we propose allows us to apply some other well-known distance measures, such as the Minkowski distance, Manhattan distance, Euclidean distance, etc., which are more precise ways of measuring similarities between phones and phonetic systems of different languages. To exemplify the difference in precision, let us look at the Hamming ($Dist_H$) and Euclidean ($Dist_E$) metrics applied to the phones [p] and [k] in the dimension 'Place of articulation' ($D_5$):

– $Dist_H = 1$: this simply means that the two phones differ in this dimension; $Dist_H$ is not able to indicate any more subtle differences
– $Dist_E = 7$: this takes into account the vector value of both phones in the dimension, which is 8 (upperlabiality) for [k] and 1 (velarity) for [p].

## 2. Computer application

Measuring the similarities between phones and articulatory systems of different languages is only one of many analytic possibilities that our framework allows. Regardless of the kind of analysis that is to be conducted within the phonetic grammar, the best results will be achieved by means of computer-aided computational methods. A mathematical model of articulation makes this approach even more plausible. Another important part of our phonetic project is to design a computer application to facilitate the research. Below is a list of results generated by those analytical algorithms which have already been implemented in the software:

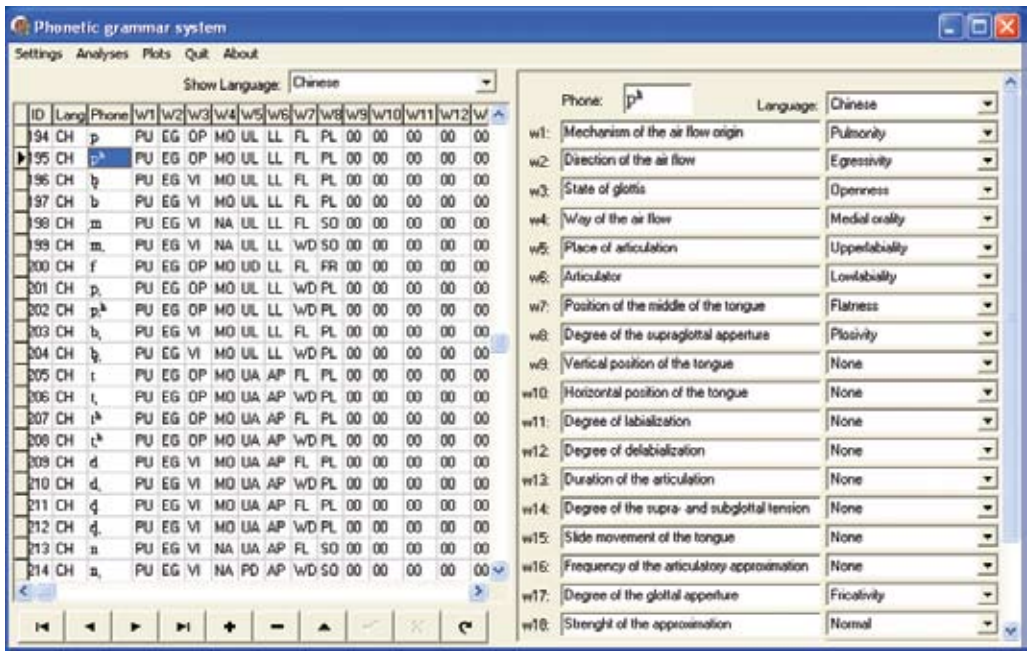
– the articulatory category[6] of a given articulatory feature (or set of features)
– the average articulatory distance between phones
– the dimensions in which given phones differ
– the combinability of a given set of articulatory features
– the most numerous articulatory category specified by a given number of features
– the number of pairs of phones being discerned by particular sets of features.

The interface of the application we have designed makes it possible:

– to input the articulatory dimensions and features occurring in them into the system
– to assign proper numerical values to the dimensions
– to define the number of languages
– to introduce a repertory of phones of a given language and a description of the phones in terms of the relevant set of articulatory features.

Below we show an example where the Chinese phone [$p^h$] is input into the application for further processing:

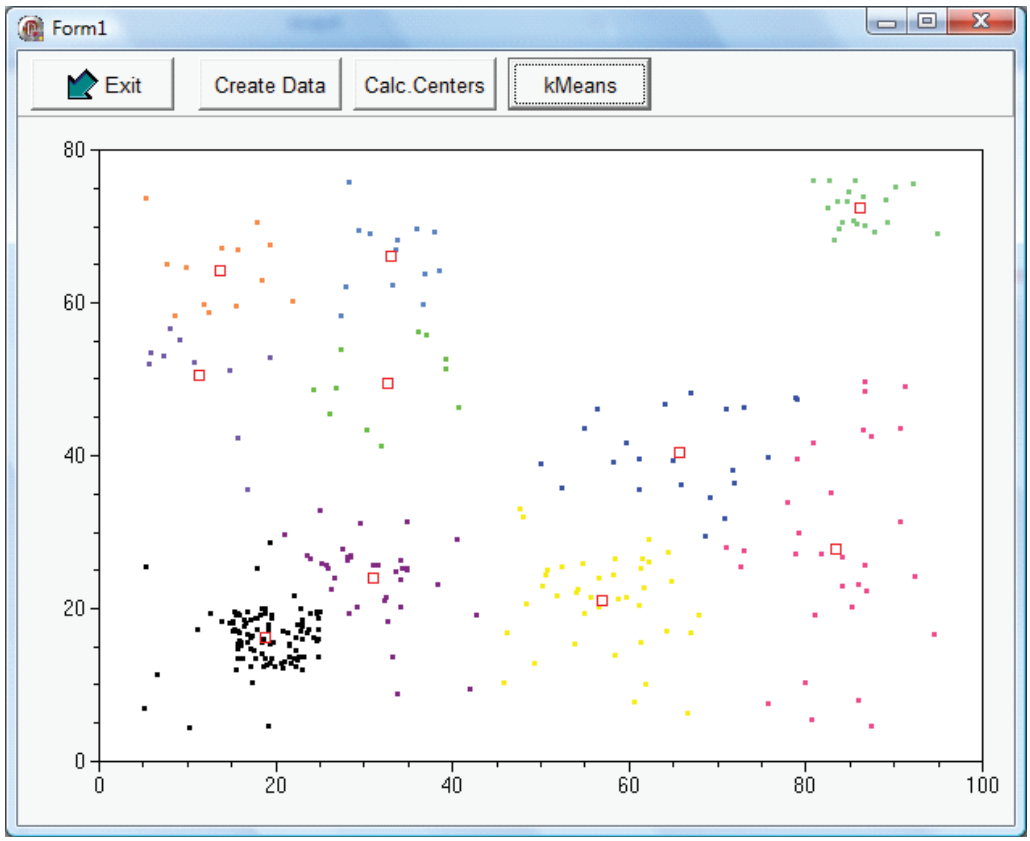
[6] Which is understood as a set of phones, each phone possessing a given feature (or features).




Data mining algorithms are used to explore the interdependencies between phones, features and dimensions. For data mining we use rudimentary methods borrowed from statistics and combinatorics. This procedure will hopefully enable us to identify certain relations between languages which have so far been unnoticed. All algorithms applied here use measures of distances as measures of similarity between phones. The algorithms can be used for phones of either one or more inventories (one or more languages).

1) K-means algorithm:[7] The first of the algorithms requires an expected number of phone clusters as an input. It makes it possible to divide the phone inventory into a particular number of disjoint classes. For example, putting $k := 2$ results in division of the set of phones into vowels and consonants. The algorithm is composed of the following steps:
   1. Place $k$ points into the space represented by the phones that are being clustered. These points represent initial group centroids.
   2. Assign each phone to the group that has the closest centroid.
   3. When all objects have been assigned, recalculate the positions of the $K$ centroids.
   4. Repeat steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the distance to be minimized can be calculated.

[7] Hand & Kamber 2000; Larose 2005.




2) The connected subgraphs algorithm:[8] This algorithm does not require the number of clusters to be input. It finds them itself on the basis of regularities in the data. The algorithm operates on the basis of the matrix of distance *DM*. This is a symmetric matrix *N x N* (where *N* indicates the number of phones taken into account) in which on the intersection of the columns and rows one obtains the distance between the relevant pair of phones, and the diagonal entries are zero. The algorithm comprises the following steps:

1. The distance matrix *DM* is calculated using a fixed distance.
2. Below the fixed threshold $\alpha$ all values in the matrix *DM* are zeroed. Finding the threshold is the basic element of the algorithm. In the simplest case it can be fixed as an average distance in the phone inventory reduced by the standard deviation of an average distance. The choice of the proper threshold mirrors our understanding of how great the distance between phones must be for them to be considered too distant to be members of the same group.

[8] Brandes et. al. 2003; van Dongen 2000.

3. Such a matrix is treated as a directed weighted graph in which non-zero values will mark weighted edges between the phones as vertices (also called a threshold graph).
4. The Depth-First Search (DFS) algorithm is applied. This algorithm results in the finding of connected subgraphs. The subgraphs are the desired clusters.

3) Agglomerative hierarchical clustering and dendrograms:[9] An agglomerative hierarchical clustering procedure produces a series of partitions of the data, $P_k$, $P_{k-1}$,…, $P_1$. The first, $P_k$, consists of $k$ single phone clusters, and the last, $P_1$, consists of a single group containing all $k$ phones. At each stage the method joins together the two clusters which are closest together (most similar). At the first stage, this amounts to joining together the two objects that are closest together, since at the initial stage each cluster has one phone.

There are several different methods that use different ways of defining distance (or similarity) between clusters:

1. Single linkage clustering: This is one of the simplest agglomerative hierarchical clustering methods. It is also known as the nearest neighbor technique. The defining feature of the method is that the distance between groups is defined as the distance between the closest pair of objects, where only pairs consisting of one object from each group are considered.
2. Complete linkage clustering: Also called the farthest neighbor, this clustering method is the opposite of single linkage. Distance between groups is now defined as the distance between the most distant pair of objects, one from each group.
3. Average linkage clustering: Here the distance between two clusters is defined as the average of distances between all pairs of objects, where each pair is made up of one object from each group.
4. Average group linkage: With this method, groups once formed are represented by their mean values for each variable, that is, their mean vector, and inter-group distance is now defined in terms of the distance between two such mean vectors.

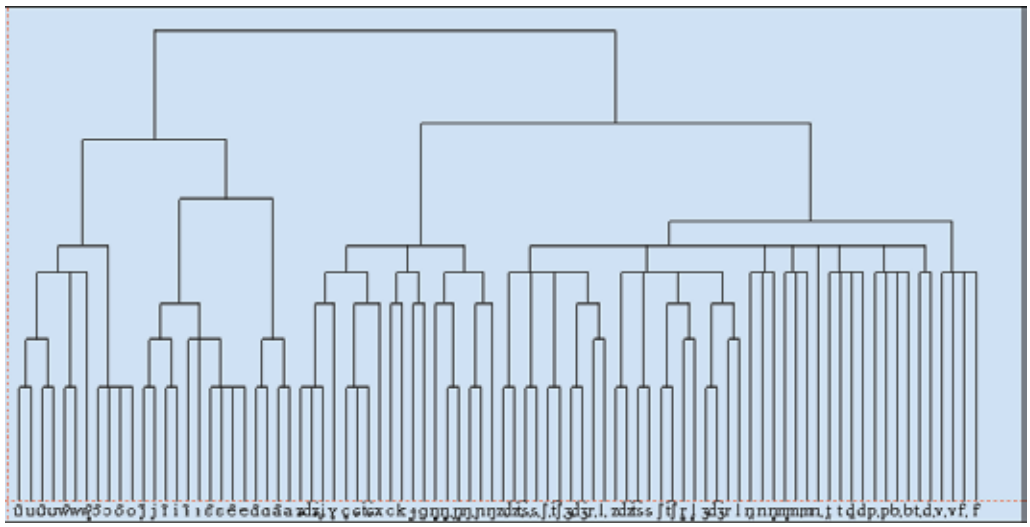
The result of the operation of the algorithm is presented in the dendrogram. This is a special type of dendric structure which provides an easy way of presenting the results of a hierarchical grouping. Cutting the dendrogram at the selected level, we can obtain a proper division into a particular number of groups of phones.

The example below is a result of application of the algorithm to the inventory of Polish phones.[10]

---

[9] Hand et. al. 2001; Jain & Dubes 1988.

[10] Mandarin Chinese inventory is not ready for this type of operation at the current stage of research.



## 3. Phone inventory of Mandarin Chinese (MC)

Every ethnic language displays language-specific problems with regard to the construction of its phone inventory. Despite the phonetic diversity of languages, in order to keep a comparative analysis viable, the homogeneity of the criteria used to build phonetic inventories must be maintained. Chinese is no different – in the following section we will discuss some specific problems and suggest solutions, although some questions will remain unsolved.

Earlier we postulated that articulatory phonetics and phonology should be kept separate, at least at the very basic level of anatomical description of the process of articulation. This postulate is crucial in our framework, but its application to Mandarin Chinese requires some extra effort – there are no existing ready-to-use inventories of Chinese phones that would suit our purposes. This is mainly because their shape is a result of a strong influence of phonological considerations. In other words, such inventories are designed by default to include a minimal number of elements (phones) to reflect phonological properties of the Chinese system. In the worst case the resulting inventory is a list of phonemes.[11] Two prominent studies on MC phonetics may serve as an example. C h a o Yuanren in his still referential grammar includes 23 consonants in the "initials" inventory, among which [b̥], [d̥] and [g̊] are listed as the only unaspirated stops.[12] Earlier in the

[11] To show the degree of accuracy of description and degree of independence of phonetic properties that are required, we shall avail ourselves of the example of a Polish inventory, which in terms of the criteria used for its construction we consider a template. The Polish phone inventory has 86 elements – 25 vowels and 61 consonants. Using the same criteria we found 21 vowels and 86 consonants in Hindi.

[12] C h a o 1968: 22.

text Chao states clearly that [b̥], [d̥], [g̊], which are lenis in nature, result "in true voiced [b, d, g] in intervocalic positions".[13] This remark is not reflected in his inventory – it is obvious that distributional variants of phonologically indistinguishable phones are not taken into account. On the other hand Chao was not really attempting to build an inventory of phones in our sense, and he is clear about his phonological priorities. Duanmu's (2000) treatment is an even better example of why it is impossible to apply existing phonetic descriptions of MC for our purposes. Duanmu clearly separates phonetics from phonology and constructs a "sound inventory" of MC, but in explaining the principles for deciding of how many sounds there are in a language, he explicitly uses certain phonological criteria like minimal pair and complementary distribution.[14] As a result he posits an inventory consisting of 59 elements: 53 consonants, 5 vowels and zero onset.[15] It is not our aim to provide a detailed commentary on Duanmu's proposal, but we will restrict ourselves to the same problem we identified in Chao's analysis – the unaspirated stops. Duanmu also lists three of these: [p], [t], [k]. The whole inventory is different from his phoneme inventory, so the criteria are not purely phonological, but they are not purely phonetic either: "…all obstruents (stops, affricates and fricatives) in Standard Chinese are unvoiced. …The unaspirated stops and affricates [p t k t͡s t͡ʃ] can become voiced [b d g d͡z d͡ʒ] when they occur in an unstressed syllable…".[16] What Duanmu does then, as is typical for phonetic studies of MC, is to apply some arbitrary decisions that are based on knowledge of phonology at a very basic phonetic level. This results in disregarding some data which are valuable for our purposes.

There are a few very fundamental theoretical commitments to be made in order to construct the MC phones inventory.

1. Decide whether to include voiced obstruents. This problem was illustrated above; we have decided to include voiced obstruents.
2. Another problem with Chinese obstruents is the treatment of their lenis nature. There is a rather puzzling discrepancy in descriptions of the unaspirated stops in literature. In the previous section we demonstrated Chao's and Duanmu's treatment – both showed little concern for delving deeper into the matter, the former mentioning the lenis nature of unaspirated stops, the latter simply neglecting the voiced stops altogether. Most other phonetic descriptions are also rather negligent in this respect. For example, from Huang's description we learn that the phone [p] should be more precisely transcribed as [b̥] – the lenis counterpart of the voiced bilabial stop.[17] The same description is applied to all unaspirated stops. The problem is that the lenis nature of a phone does not mean that it is voiceless. On the other hand the lack of phonological distinction between voiced and voiceless phones in MC has no significance for the phonetic grammar. The question to be answered is what and

[13] Chao 1968: p. 21.
[14] Duanmu 2000: 12–15.
[15] Duanmu 2000: 44.
[16] Duanmu 2000: 26–27.
[17] Huang et al. 2008: 116.

how many phonetically distinct unaspirated stops there are in MC. There are several theoretical possibilities involving voiceless, voiced and lenis stops. The bilabial series will serve as an example: [p b̥ b]. Our preliminary treatment is to include all three phones in the inventory. In order to be able to do that, a new articulatory dimension had to be added to the list – $D_{18}$: the strength of approximation – with the set of features {none, weak, normal}. This allows us to introduce the opposition lenis: non-lenis in addition to the existing voiced:voiceless. The lenis phones are assigned the *weak* feature.

3. There is no agreement in respect of the treatment of the consonant + glide combinations in MC. Most studies analyze such combinations as two sounds, but some prominent ones (i.e. Chao 1968, Duanmu 2000) argue that they should be treated as one sound. The latter treatment adds 29 phones to the inventory. For our purposes we have decided to adopt an intermediary solution. Duanmu's arguments in favor of the single sound treatment are phonological in nature. Phonetically it is less viable to consider the C+G combinations as one sound. Treating them as two sounds instead of one should not cause the loss of any important articulatory characteristics of MC, which would be an important argument against this solution. For this reason we cannot totally disregard the C+G combinations, because in some cases that would lead to the information loss. As a general rule we treat the C+G combinations as separate sounds, except for the cases where there is no alternative to the single-sound treatment. This is the case with those C+G combinations where dental consonants are not replaced by palatals in some variants of Chinese pronunciation.[18] The dental+glide combinations are included in our inventory, being identified either by the feature *semipalatality* or by *semilabiality*. In this approach we obtain a set of 8 C+G combinations in the inventory of MC, instead of the 35 of Duanmu's proposal: [t͡sʲ d͡zʲ t͡sʰʲ sʲ t͡sᶣ d͡zᶣ t͡sʰᶣ sᶣ]. Another result of our theoretical commitments is the inclusion of two voiced dentals in this series. This approach to the consonant-glide combinations should be considered a preliminary one and a subject to change.
4. We include all possible distributional variants of phones, e.g., palatalized consonants.

As a result we constructed the following inventory of Mandarin Chinese phones:

[p pʰ b̥ b pˌ bˌ pˌʰ b̥ˌ t tʰ d̥ d tˌ dˌ tˌʰ d̥ˌ k kʰ g g̊ m mˌ n nˌ ɲ ŋ ŋˌ l lˌ r f s ɕ ʃ ʒ x t͡s t͡sʰ d͡z d͡z̥ t͡ɕ t͡ɕʰ d͡ʑ d͡ʑ̊ t͡ʃ t͡ʃʰ d͡ʒ d͡ʒ̊ t͡sʲ d͡zʲ d͡z̥ʲ t͡sʰʲ sʲ t͡sᶣ d͡zᶣ d͡z̥ᶣ t͡sʰᶣ sᶣ j ɥ w i y u o E ɤ e ə ʌ ɒ a æ ɐ].

A detailed articulatory description of every phone, as well as the results of applying the computational analysis introduced in section 2., are beyond the scope of this paper.

[18] Duanmu describes this type of pronunciation as common for women and children (Duanmu 2000: 33).

## References


Bańczerowski, J. 1985, "Phonetic relations in the perspective of phonetic dimensions", In: Pieper U., Stickel G. (eds.) 1985. *Studia Linguistica Diachronica et Synchronica*. Mouton de Gruyter.

Bańczerowski, J. 1987, "Towards a dynamic approach to phonological space", *Studia Phonetica Posnaniensia* vol. 1, 5–30.

Bańczerowski, J. 1990, "Undular aspect of phonological space-time", *Studia Phonetica Posnaniensia* vol. 2, 13–42.

Bańczerowski, J. 1992, "Formal properties of neostructural phonology", *Studia Phonetica Posnaniensia* vol. 3, 5–28.

Bańczerowski, J., Pogonowski J., Zgółka T. 1982, *Wstęp do językoznawstwa*. UAM, Poznań.

Batóg, T., 1967, *The Axiomatic Method in Phonology*, Routledge and Kegan Paul.

Brandes, U., Gaertler, M., Wagner, D. 2003. "Experiments on graph clustering algorithms", *Lecture Notes in Computer Science, Di Battista and U. Zwick (Eds.)*: 568–579.

Chao, Yuanren. 1968. *A Grammar of Spoken Chinese*, University of California Press.

van Dongen, S. 2000. *Graph Clustering by Flow Simulation*, PhD thesis, University of Utrecht.

Duanmu, San. 2000. *The Phonology of Standard Chinese,* Oxford University Press.

Grimes, J.E., Agard, E.B. 1959, "Linguistic divergence in Romance", *Language* 35: 598–604.

Hand, J., Kamber, M. 2000. *Data Mining: Concepts and Techniques,* Morgan Kaufman.

Hand, J., Mannila, H., Smyth, P. 2001. *Principles of Data Mining*, MIT Press.

黃錦鋐 (Huang Jinhong), 張正男 (Zhang Zhengnan), 張孝裕 (Zhang Xiaoyu), 黃家定 (Huang Jiading), 葉德明 (Ye Deming) 2008. 國音學 (Mandarin Chinese Phonetics), 中正書局, 臺北.

Jain, A., Dubes, R. 1988. "Algorithms for Clustering Data", Prentice-Hall.

Larose, D.T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*, Wiley.

Meyer-Eppler, W. 1963. *Grundlagen und Anwendungen Informationstheorie*, Springer Verlag.

Steffen-Batóg, M. 1997, *Studies in Phonetic Algorithms*, Poznań: Sorus.

Steffen-Batóg, M., Batóg, T. 1980, "A Distance Function in Phonetics", *Lingua Posnaniensis* 23, 47–58.