PHONOLOGY AS HUMAN BEHAVIOUR:
CLINICAL PHONETICS, PHONOLOGY AND PROSODY

YISHAI TOBIN
Ben-Gurion University of the Negev, Be’er Sheva
yishai@bgu.ac.il

ABSTRACT

This paper introduces the theory of Phonology as Human Behaviour (PHB); summarises the basic theoretical and methodological tenets of the theory and shows how it has been applied to clinical phonetics, phonology and prosody. The theory of PHB, developed by William Diver and his students of the Columbia School, combines aspects of the “communication factor” inherent in Prague School phonology with aspects of the “human factor” inherent in André Martinet’s functional diachronic phonology. The major parameters of the theory are presented according to the Saussurean-based semiotic definition of language as a sign system used by human beings to communicate. The fundamental axiom underlying the theory is that language represents a compromise in the struggle to achieve maximum communication with minimal effort. The major contribution of the theory is that it provides a motivation to explain the non-random distribution of phonemes within the speech signal in language in general and in typical and atypical speech in particular.

KEYWORDS: PHB; clinical phonetics; phonology; prosody.

1. Introduction

This paper (i) introduces the theory of Phonology as Human Behaviour (PHB) (also known as Columbia School (CS) Phonology); (ii) summarises the basic theoretical and methodological tenets of the theory and (iii) shows how it has been applied to clinical phonetics, phonology and prosody. The theory of PHB adheres to a basic functional and semiotic definition of language as a sign system used by human beings to communicate. The fundamental axiom underlying the theory is that language represents a compromise in the struggle to achieve maximum communication by expending minimal effort. The major contribution of the theory is that it provides a motivation to explain the non-random distribution of phonemes in the structure of language per se as well as within the speech signal in both typical and atypical speech.
The theory of PHB was introduced by William Diver (1979) in an analysis of the non-random distribution of certain classes of initial consonant clusters in English. The specific initial consonant clusters studied were of the type $C_1C_2$ where $C_1$ is either a mobile phoneme (e.g. a stop which requires the movement of the active articulators for closure and release) or a stable phoneme (e.g. a fricative which holds the active articulators in a steady position) and $C_2$ is either a mobile /r/ (originally trilled) or a stable /l/ (lateral approximant). A similar study followed for initial consonant clusters of this type for Italian and Latin (Davis 1987). These studies discovered and supported the following phonological principles in an ascending order of strength:

1. like phonemes, i.e. phonemes made with similar types of articulatory gestures, prefer to collocate with each other (sames are favoured, differents are disfavoured): mobile + mobile and stable + stable;

2. additional sets of articulators and articulatory gestures are less favoured: voiceless versus voiced;

3. co-articulation by the same articulators/phoneme is even more highly disfavoured.

These three principles regarding the non-random distribution of initial consonant clusters of this type based on the theory of PHB have been further tested and supported for over forty-two different languages representing nine diverse language families [Indo-European: Germanic (German (Middle, High/Modern), Yiddish, Dutch, Afrikaans, Swedish, Norwegian, Danish); Romance (Latin (Vulgar/Classical), French, Spanish, Portuguese, Catalan, Romansch, Sardinian, Romanian); Celtic (Irish, Welsh); Slavic (Russian, Ukrainian, Polish, Czech, Slovak, Serbo-Croatian, Bulgarian); Baltic (Lithuanian, Latvian); Hellenic (Classical/Modern Greek); Albanian; Indo-Iranian (Bukharian (Judeo-Persian))]; [Semitic (Hebrew, Arabic (Classical/Modern, Amaraic)); [Finnno-Ugric (Finnish, Estonian, Hungarian)]; [Caucasian (Georgian (Grusinski))]; [Dravidian (Malayalam)]; [Tibetan]; and [“artificial” languages (Esperanto, Klingon)] (Tobin 1997a, 2000a). Similar results for the phonotactic distribution of initial consonant clusters of the type $C_1C_2$ where $C_1$ = a mobile/stable phoneme of constriction and $C_2$ = /r/ or /l/ were found for all the languages studied above. The same favourings and disfavourings supporting principles 1–3 were obtained across languages albeit with slightly different degrees of compatibility which may be influenced by diachronic and extra-linguistic factors.

Diver (1993, 1995) later expanded his study of initial consonant clusters to explain the non-random combinations of vowels and consonants in English and in language in general. Subsequently, the theory of PHB was further extended to explain the combina-
tory phonology of consonant and vowel phonemes in a large number of diverse languages from several different language families, e.g. Latin, Italian (Davis 1987); Hebrew (Tobin 1990b–c); Urdu (Azim 1989, 1993, 1995, 1997; Hameed 2004, Jabeen 1993); Mewati (Fatih 1987); Spanish (Flores 1997; Dekker and de Jonge 2006); and Byelorussian (Dreer 2006).

As will be seen in this paper, PHB has also been applied to the areas of developmental and clinical phonetics, phonology and prosody for a wide variety of languages, e.g. English (Tobin 1997a,b, 1999, 2002); Spanish (Enbe 2009; Enbe et al. 2006); Hindi-Urdu (Fatih 2007); Japanese (Miyakoda 2003, 2004a,b; Tobin and Miyakoda 2004a,b, 2006); Finnish (Moore 1991a,b, 1993; Moore and Korpiaakko-Huukka 1996; Moore and Rosenberg-Wolf 1998); Polish (Polezyńska 2009; Polezyńska and Tobin 2009); Israeli Hebrew (Green 2009; Green and Tobin 2008a,b, in press; Halpern and Tobin 2008; Paltiel-Gedalyovich 2009; Tobin 1995, 1997a, 2000a,b; Zivan 2009); and Israeli Sign Language (ISL) (Fuks 2009; Fuks and Tobin 2008, 2009; Tobin 2007a,b, 2008). Moreover, PHB has been combined with other phonological theories such as Natural Phonology (NP) and Optimality Theory (OT) in additional developmental and clinical analyses of Hebrew (e.g. Adi-Ben-Said 2006; Adi-Ben-Said and Tubul-Lavi 2009; Ben-David 2001); Polish (e.g. Polezyńska-Fiszer 2006, this volume); and Jordanian Arabic (Bader, this volume; Bader and Gibbon 2008).

Other areas of research to which PHB has been applied include: the interfaces between (a) diachronic combinatorial phonology and inflectional and derivational morphology across languages, e.g. Old, Middle and Modern English (Tobin 2006); Old Church Slavonic, Old and Modern Russian (Buk 2003); Biblical and Modern Hebrew (Perelstein 2008); Classical, Modern and Urban Palestinian Arabic (Saif 2004); Hungarian (Salmon 2003); Latin (Cohen 2001); Latin, Spanish and Portuguese (Oron 2003); Ladino (Tubul 2002); and (b) phonology and the lexicon in the Hebrew triconsonantal (CCC) root system (Tobin 2004; Nissan 2007); in a literary text (Podzhrebin 2005; Roe-Portianski 2007); and in the definite article in Yemenite Arabic (Ali and Hameed 2007). PHB has also been applied to a wide range of other historical, psycholinguistic, and sociolinguistic issues in various languages (e.g. Contini-Morava, Kirsner, and Bachiller-Rodriguez 2004; Davis et al. 2006; Enbe and Tobin 2007; Joue and Collier 2006; Reid et al. 2000; Tobin 1993, 2001, 2004). Furthermore, recent studies have also applied and implemented PHB to a computer program based on lip-reading for the hearing impaired (Schocken 2008; Schocken et al. 2008). Finally, PHB has been compared to and contrasted with other functional, formal and quantitative-oriented phonological theories such as Phonometrics (e.g. Zwirner and Zwirner 1966, 1970; in Tobin 1988c); Prague School phonology (e.g. Tobin 1988a, 1997a–c, 2007d); Natural Phonology (NP) (e.g. Donegan and Stampe 1979; Dressler 1996; Dressler et al. 1987; Dziubalska-Kolaczky 2001, 2002; Dziubalska-Kolaczyk and Weckwerth 2003; Stampe 1972/1979; in Tobin 2007c, 2009); and Optimality Theory (OT) (e.g. Archangeli and Langendoen 1997; Kager 1999; McCarthy 2002, 2003, 2008; in Tobin 2000b. 2009).
3. Placing PHB in its historical framework

PHB may be viewed as part of the historical development of a larger twentieth-century structural, functional, cognitive and naturalistic approach to linguistics and phonology. This tradition begins with Ferdinand de Saussure’s (1916/1972) semiotic concept of system and the dichotomies of *langue* and *parole*, and phonetics-versus-phonology, based on a classification of sounds according to their articulatory and acoustic features. This fundamental dichotomy between the abstract code and its concrete realisation based on the concept of distinctive features that create communication oppositions was further developed by Nikolai Trubetzkoy (1939) and Roman Jakobson (1941) of the communication-oriented Prague School (PS) (Tobin 1988a, 1997, 2007d). The strict opposition-based communication factor adhered to by the PS was further supplemented by the introduction of the human factor to phonology through the concepts of “asymmetry” and “economy of effort in phonological change” by André Martinet (1955). Martinet maintained that phonological systems are arranged asymmetrically and change in such a way that their non-random diachronic distribution reflects the search for equilibrium and harmony within the system as it is affected by the principle of least effort in human behaviour. This principle of minimal effort postulated by Martinet implies that speakers strive for a minimal number of distinct phonemes which requires the least amount of effort to be produced and combined together in what Sampson (1980) reviewed in Tobin (1986) referred to as the “therapeutic view of sound change”.

It is Diver (1975, 1979, 1995), however, who has shown that a more complete theory of phonology has to take both the communication factor and the human factor into account. Diver maintains that there is a constant struggle between our need for maximum communication and our desire for minimum effort. The communication factor (requiring a large number of perceptually distinct phonemes demanding a great deal of effort) is in conflict with the human factor (striving for minimal effort), which results in a trade-off between the two. This synergetic compromise between the communication factor and the human factor is reflected in the fact that there is a similar number (20–40) of phonemes of varied proportional degrees of difficulty acquired in a similar order in the languages of the world: less than 20 phonemes might reduce the communication potential and more than 40 might become rather difficult to learn, remember and produce. Thus, Diver extends and enhances Martinet’s more strictly diachronic view of the human factor so that it can become a means of explaining the non-random distribution of phonemes in language in coordination with the needs of communication as originally established by Saussure and the PS (discussed in Liberman 1991; Tobin 1988a-c, 1990a, 1997a, 1997d, 2007d).

3.1. Defining language semiotically

PHB represents the phonological component of a larger linguistic theory originally called Form-Content Analysis (FCA) and now referred to as the Columbia School (CS).
FCA originally defined language as: “a system of systems composed of various sub-systems revolving around the notion of the linguistic sign which are organised internally and systematically related to each other and used by human beings to communicate” (Tobin 1990a, 1993, 1994). This definition has been subsequently simplified to: “language is seen as a symbolic tool whose structure is shaped both by its communicative function and by the characteristics of its users” (Tobin 2007a–b, 2008).

3.2. Defining language synergetically

The particular definitions of language and theoretical and methodological implications with regard to *langue* and *parole* mentioned above provide a holistic view of language based on the relationship between the communication factor and the human factor. Both these factors can be further related to the larger holistic concept of *synergesis*, which may be defined as the cooperative action of discrete agencies with the result that the total effect is greater than the sum of the discrete effects taken independently. Put more simply: the linguistic whole (language) is greater than the sum of its individual parts (i.e. the various units that may be related to signals and meaning and their use: phonemes, syllables, morphemes, stems, roots, words, word classes, meanings, phrases, clauses, sentences, utterances, contexts, texts, etc.). Furthermore, concerning the communication and human factors, successful linguistic communication is achieved only through the combined effort of an encoder and a decoder cooperating together (Tobin 1990a, 1993, 1994, 1997a). According to Zipf (1949), from the point of view of the encoder, linguistic economy is best realised by a highly compressed lexicon composed of exclusively short words conveying a multiplicity of messages, or what Zipf calls *the force of unification*. From the point of view of the decoder, on the other hand, linguistic economy is best realised by a lexicon composed of words maximally distinct in form and restricted to one exclusive message, or what Zipf calls *the force of diversification*.

Therefore, it is possible to claim that the semiotic act of communication can be seen as a “mini–max” struggle: the desire to create maximum communication with minimal effort. However, this mini–max struggle is reflected by two converse communication processes engaged in by encoders and decoders who stand in opposed and mirror-like positions. Minimal effort on the part of the encoder will place a heavier burden on the decoder’s inferential abilities. The opposite is also true. Minimal effort on the part of the decoder may result in a breakdown of communication that most often can be remedied by an increased effort on the part of the encoder: that is, by forcing the encoder to choose signs in a way that may be inferred with less effort on the part of the decoder. In short, this mini–max struggle for linguistic communication is controlled by the synergetic principle that the more cooperation there is between the encoder and the decoder – within the appropriate linguistic and situational contexts of the speech act – the greater the chance there is for successful communication (Tobin 1990a:59).
Thus, synergesis can be viewed as a fundamental concept underlying language and the behaviour of language use(rs), one that may be appealed to in order to explain disparate linguistic phenomena in a motivated, systematic and holistic way. Concerning phonetics versus phonology, the synergetic relation between the encoder and the decoder is found in the encoder’s production of a large number of concrete allophones in different phonetic environments (each allophone appearing in the easiest or most natural environment – the human factor) that are perceived as a more limited number of abstract phonemes by the decoder (in order to create meaningful oppositions – the communication factor; Tobin 1997a, 2007d).

4. The parameters of PHB

To reiterate: the theory of PHB combines aspects of the “communication factor” inherent in PS phonology with aspects of the “human factor” inherent in Martinet’s diachronic phonology. The major parameters of the theory are presented according to the functional semiotic definition of language as a sign system used by human beings to communicate. The fundamental axiom underlying the theory is that language represents a struggle between the need to achieve maximum communication (the communication factor) with the desire to expend minimal effort (the human factor). The major contribution of the theory of PHB is that it provides a motivation and an explanation for the distribution within the speech signal, i.e. it tells us why the distribution of phonemes within a language is not random but motivated.

4.1. The four orientations underlying the theory

Based on this definition of language and the above axioms inherent to the theory, one can list the four orientations underlying PHB as:

(1) the communication factor;
(2) the physiology of the vocal tract;
(3) the acoustic medium;
(4) the human factor.

The communication factor and the human factor will be further elaborated upon in this paper. The physiology of the vocal tract is related to how human beings learn to control specific musculature to alter the air stream in the production of phonemes (articulatory phonetics) and the acoustic medium refers to how human beings perceive these phonemes (acoustic phonetics). The orientations of PHB are directly related to the semiotic and synergetic definitions of language presented above.

Concerning the fundamental principles which can be used to define, compare and contrast linguistic and phonological theories found in the introduction of this paper: (1)
Language is defined in terms of its function as a system of communication that ecologically reflects the characteristics of its users – human beings. This particular definition of language encompasses the synergetic interaction between the communication and the human factors which determines the raison d'etre and dictates the formulation of the specific problems being addressed as well as the source, amount and manner in which the data are identified; collected and analysed; and evaluated.

The communication factor for the theory of PHB may be summarised as the systematic and non-random combination of distinctive phonological units to create meaningful oppositions in linguistic signs of all sizes and levels of abstractness (words, parts of words, expressions, zero morphology, word order signals, sentences, prosodic intonation, pauses, changes in tone, and even larger units such as texts (Tobin 1990a: 30).

The human factor includes the following basic principles concerning phonetics and phonology in theory of PHB:

1. **Human intelligence.** Human beings can draw far-reaching abstract conclusions from minimally salient concrete cues through the cognitive process of inference: we infer (i.e. perceive and are aware of) the abstract phoneme despite the fact that we hear the concrete allophone that was actually uttered.

2. **Human efficiency.** Human beings invest minimal effort for maximal results in the semiotic communication process: by inferring abstract phonemes rather than concrete allophones, we have to invest just enough to perceive only those distinctive articulatory and acoustic features necessary for communication, rather than all the primary and secondary features associated with each individual allophone.

3. **Memory limitations.** Human beings have large but limited memories that can be directly related to human intelligence and efficiency. It is easier to remember 20–40 distinctive phonemes that are used to create meaningful oppositions rather than a much larger number of allophones that do not produce meaningful oppositions.

One of the aspects of the theory of PHB that separates it from many other phonological theories is its recognition of the interdependence of the abstract phonological system of langue and the concrete distribution of sounds in parole. This integral connection between phonetics and phonology is supported by a theoretical and methodological model based on sound units similar to what have traditionally been called phonemes versus allophones (cf. Diver 1995: 62–72; Davis 2006). These phonemes and allophones are defined as linguistic units containing various features which systematically link them together according to their communicative function and reflect the characteristics of human perception, cognition and behaviour relevant to their exploitation (Tobin 1997a: 19–23). This integral connection between phonetics and phonology further implies a
synergetic and holistic search for basic principles that can account not only for the
sounds of language, but for all aspects of language as well, based on the interaction be-
tween the communication and human factors. Therefore it is not surprising that the
same principles underlying PHB can be integrated into the analysing of sounds of lan-
guage per se, i.e. viewing articulatory, acoustic and distributional phonetics from the
point of view of human behaviour from both the point of view of the encoder and the
decoder in accordance with the linguistic and situational contexts.

5. Viewing phonetics/phonology as human behaviour

The following principles underlie the articulatory and acoustic phonetic implications of
PHB. Articulatory phonetics in general is viewed as the history of the air stream from
the diaphragm to the release of the air. This history of the air stream entails a process
whereby inhaled air is expelled from the lungs, transformed into acoustic energy (voice)
by the action of the vocal folds, modified (filtered) by the dynamic changes in the con-
figuration of vocal tract, through the movements of the articulators, and radiated from
the oral or nasal passages by the encoder. Each of the articulating organs is viewed from
the point of view of its contribution to systematically modifying the airstream to create
new acoustic patterns which are perceived as different sounds by the decoder. Acoustic
phonetics therefore may be viewed as the history of the airstream from its release by the
encoder to its perception by the decoder. The different acoustic patterns produced by the
encoder and perceived by the decoder are analysed according to the distinctive articulat-
ory features produced by the encoder to each of which there is a parallel acoustic dis-

tinctive feature. Both the production and the perception aspects of human speech are
viewed together as part of the same integral process referred to as the speech chain (De-
nes and Dinson 1963).

The following principles underlie the classification of sounds or phones into pho-
nemes and allophones. Phones are classified as phonemes versus allophones based on: (1)
their contribution to communication (phoneme versus allophone): the communication fac-
tor; (2) the specific primary distinctive versus secondary non-distinctive articulatory and
acoustic features which speakers must learn: the human factor; (3) the relative non-
predictability or openness of the distribution of phonemes in minimal pairs (the commu-
nication factor) versus the predictability of the complementary distribution of allophones
(the human factor). There is a systematic interchangeability of phonemes and allophones
across languages and within the same language in different periods of time that can be ex-
plained from the points of view of the communication and human factors. The fact that all
languages have a similar number of phonemes (usually between 20–40) which are ac-
quired in a similar order across languages and their diachronic versus synchronic relation-
ships to alphabet systems and orthography can be related to both factors. The asymmetric
relationship between the number of phonemes versus the number of allophones in lan-
guage systems and our awareness of phonemes versus allophones can be explained from
the synergetic point of view of achieving maximum communication through minimal effort (Tobin 1997a: 18–23).

The following principles derived from the studies of a large number of languages from different language families in the PHB framework underlie the replacement of traditional and neo-traditional phonetic categories and labels with new concepts which illustrate the unity of sounds and sound classes as part of a single hierarchy (cf. Davis 1987; Tobin 1997a). Research in PHB has uncovered and supported a set of specific phonological and phonotactic parameters explicitly derived from the theory such as:

(1) Consonants are referred to as **phonemes of constriction**.

(2) Vowels are referred to as **phonemes of aperture**.

(3) Features such as voiced versus voiceless, nasal versus oral, emphatic versus non-emphatic, ejective versus non-ejective, retroflexed versus non-retroflexed, palatalised versus non-palatalised, and labialised versus non-labialised, etc., are studied in terms of the number of articulators being exploited simultaneously and the degree of the difficulty of their control from the point of view of the human factor.

(4) Place of articulation for phonemes of constriction is replaced with the notions of **active articulators** versus **passive receptors**.

(5) The tongue and the lips are viewed as active articulators for phonemes of aperture concerning the height and position of the former and the degree of rounding or spreading for the latter.

(6) The division of the tongue into **anterodorsum** and **posterodorsum** as active articulators does not only serve to distinguish phonemes of constriction but can also replace the traditional categories of front versus back vowels respectively as the active articulators for phonemes of aperture.

(7) Most (if not all) of the other various traditional categories of manner of articulation of consonants and vowels versus semi-vowels such as plosives, trills, flaps, glides, liquids and approximants, etc., are replaced by a unified hierarchical set of **degree of stricture**, **aperture** and **type of airflow** that is applicable to both phonemes of constriction and phonemes of aperture which now may be viewed as a single, holistic continuum or cline of sounds.

(8) **Mobile** phonemes (e.g. stops which requires the movement of the active articulators for closure and release or a trilled /r/) versus **stable** phonemes (e.g. a fricative which holds the active articulators in a steady position or the lateral approximant /l/) can replace former categories of manner of articulation for pho-
nemes of constriction and can be applied to both phonemes of constriction and phonemes of aperture replacing categorisations such as obstruents versus sonorants and syllabic versus non-syllabic phonemes.

(9) Tenseness versus laxness and fortis versus lenis can be viewed from the point of view of the amount of effort needed to control different sets of articulators and musculature for all sounds regardless of their classification as phonemes of constriction or aperture.

(10) The difference between simple phonemes versus complex phonemes (stops and fricatives versus affricates and monophthongs versus diphthongs) can be explained from the point of view of the synergetic interaction between the desire for maximum communication with minimal effort for both phonemes of constriction and aperture in a similar way.

The following principles can explain the non-random distribution of sounds both within the phonemic inventory of a language system as well as within meaningful units of language based on the principles of PHB.

(1) Maximum communication is achieved by investing human effort in the exploitation of the extremities of the oral vocal tract which provide the most distinct acoustic cues (e.g., grave and acute loci) in opposition to a neutralised center:
   (a) the lips versus the postero-dorsum in opposition to the apex (front-back-middle) the most adroit of the articulators for phonemes of constriction;
   (b) the vowel triangle: /i/ versus /u/ in opposition to /a/ (front-back-middle) for phonemes of aperture.

(2) When these optimal oral oppositions are exploited, further communicative distinctions are made by exploiting the musculature between the extremes and the middle or by extending their boundaries:
   (a) by adding additional passive receptors (teeth, hard palate) or less adroit active articulators (antero-dorsum, uvula, pharynx, glottis) or the vocal folds (for the glottal stop /ʔ/ and fricative /h/ as well as for tone, murmur, creakiness, etc. in addition to voicing) for phonemes of constriction and aperture;
   (b) by exploiting the intermediary tongue positions between the high and low front and back points of the vowel triangle /ɛ/, /æ/, /ɔ/, /o/, etc.

(3) Additional communicative distinctions can be made by exploiting the adroit active articulators in different ways requiring more effort and thus creating marked sounds such as lateral, retroflex, emphatic, ejective, velarised, nasalised, etc. phonemes.
(4) More complex sounds requiring greater effort (affricates, diphthongs, etc.) can create further communicative distinctions as well.

(5) In all of the above processes there also will be a favouring of adroit active articulators and a preference for the excitation of fewer sets of articulators when possible as a result of the human factor.

(6) The differences in the higher communicative force of utterance-initial versus the lower communicative force of utterance-medial versus utterance-final positions also will affect the choice of more adroit versus less adroit, or more visual versus less visual articulators, and phonemes requiring one (voiceless), two (voiced), or three (nasal) sets of articulators.

(7) In most, if not all languages, the relative number and the proportion of marked versus unmarked phonemes will be similar, although the features being marked will differ from language to language.

(8) In most, if not all languages, the number and kind of phonemes and their role in syllable structure will be based on the differences of the communicative forces of phonemes of constriction which impede the airflow in relation to phonemes of aperture which provide free movement of air.

To summarise: The common denominator underlying PHB is to analyse the sound systems of language and the non-random distribution of sounds in language from the synergetic point of view of achieving maximum communication with minimal effort taking into account the different roles of encoders and decoders needed to produce efficient communication.

6. The fundamental analytic position of PHB

The motivation for the explanation of the non-random phonological distribution of phonemes within language central to the theory is based on the following seven theoretical and methodological assumptions (and their clinical applications):

(1) Users of a language behave as though they have acquired certain distinctive units – the phonemes – which they deploy for communicative purposes (analysing and distinguishing between speech errors that are phonetic versus phonemic/phonological).

(2) We cannot directly observe these distinctive units – the phonemes – that users of language deploy for communicative purposes (analysing and comparing the phonological systems and speech output of typical and atypical speakers).
We can, however, observe the non-random appearance and combinations of these phonemes – their phonotactic skewing – a skewing that has been built up over the centuries and millenia in the very mouths of the speakers (and should be the same for all speakers of the same language).

We can infer that these long-range skewings of phonemes represent favourings and disfavours on the part of users of the language. (It is to be observed that the skewings are not idiosyncratic to particular languages; their general characteristics recur from language to language. This is reminiscent of terms such as “vowel goodness” or “prototypical vowels” of a language used in the theoretical and clinical literature.)

We can then examine the favourings and disfavours against the background of the orientation – which means with independent knowledge of what kinds of favourings and disfavours humans are prone to in areas other than the use of language (control of fine motor movement in speakers in general and with other possibly related motor problems, e.g. apraxia/dyspraxia, Down’s syndrome, for atypical speakers, etc).

We can infer that a disfavouring, for example, represents a difficulty in a learning process, and by a close examination of what it is that constitutes the difficulty in that particular learning process, we can infer what it is that is being learned (analysing the difficulties of atypical speakers).

What it is that is being learned we may identify as a characteristic of the distinctive units that human beings acquire and deploy for communicative purposes. (This notion of favouring versus disfavouring is reminiscent of the tendency of infants to come into the world with the potential ability to perceive all the phonemes of all the languages of the world, but to learn to ignore sounds that are irrelevant to their own language.)

These basic theoretical and methodological assumptions can not only be applied to the non-random phonological distribution of phonemes within language and languages in general but also to both the typical and atypical speech output of individuals in particular.

7. PHB in the speech and hearing clinic

Natural Phonology (NP) (Stampe 1972/1979) is one of the most predominantly used theories in developmental and clinical phonology. NP developed at a time when linguists began to question the way naturalness was handled within the generative frame-
work. In developing his theory of natural phonology, Stampe started with the observation that the generative view of phonological rules as complexities that have to be acquired does not fit the facts of language acquisition. For example, devoicing of word-final obstruents occurs in German (or Russian), but not in English. Within generative grammar, German children are considered to “acquire” the devoicing rule. This, however, goes against what actually is observed, for children do not “learn” devoicing, rather, they “never learn” to pronounce “voiced” consonants in final position. Children whose native tongue is English commonly go through a stage where they devoice word-final consonants, but eventually “suppress” this tendency later on. It is apparent that in both cases, children start out with a tendency to devoice in final position. This fact could not be adequately accounted for within the generative framework. In NP, it is assumed that all children start out with the phonological process of devoicing, and as they learn to pronounce final voiced consonants, they “suppress” this tendency by eliminating the process of devoicing (in the case of German- and Russian-speaking children, they do not “learn” to pronounce final voiced consonants; hence, the word-final obstruent is left devoiced).

This way of capturing phonological processes led to the integration of the concept of naturalness into phonological theory. Naturalness reflects what actually is observed in children’s speech. However, the labeling, classification or categorisation of the phonological processes may describe the articulatory and phonotactic phenomena, but do not necessarily “explain” them explicitly. PHB goes one step further: the principles of PHB can explicitly account for “why” these processes occur. For example, PHB claims that devoicing (using fewer sets of articulators – the human factor) generally takes place in final position because the demands of communication are lower in final position than they are in initial position (the communication factor). In this section some of the major natural functional processes introduced by Stampe (1972/1979) as they have been applied to developmental and clinical phonology for a wide variety of languages (e.g. English, Polish, Spanish, Hindi-Urdu, Japanese, Finnish, Israeli Hebrew, Palestinian and Jordanian Arabic) will be exemplified (with Israeli Hebrew examples) and further clarified and explained according to the principles of PHB. The first step in the analysis of these natural functional processes is to determine whether they are phonological or phonetic errors, i.e. whether the abstract phonemic or phonological system of langue (competence) is affected or whether it is merely a question of parole or performance.

7.1. Functional processes influencing syllable structure

(1) Final consonant deletion: CVC → CV (chronology 2:0 → 3:2):
word: sipur → sipú ‘story’
gadol → gadó ‘big’
syllable: taxtonim → tatonim ‘underwear’
jaldá → jadá ‘girl’
syllable and word: mafiéax → matéa ‘key’ (child with Down’s Syndrome)
parpar → papá ‘butterfly’ (child with Dyspraxia)
Explanation: word final position has less communicative force; consonants require more articulatory control (are harder to make) than vowels.

(2) Deletion of unstressed syllables (chronology: 2:0 → 4:0):
   one unstressed syllable: maká → ka ‘knock’  naxón → xon ‘right’
   ambátja → bátja ‘bath’  tsaláxat → láxat ‘plate’
   two unstressed syllables: tarnególét → gólet ‘chicken’
   televisía → vízia ‘television’

Explanation: stressed syllables give more communicative, perceptual and cognitive information than unstressed syllables, the more syllables in the word, the more effort it takes to pronounce it, in non-initially stressed words, the less information in word initial position which usually has the greatest communicative force.

(3) Cluster reduction: CC → C (chronology 2:0 → 3:6-8):
   initial: praxím → paxím ‘flowers’  bgadím → gadím ‘clothes’
   medial: axálti/axálnu → axáti/axánu ‘I/we ate’ (child with Down’s Syndrome)

Explanation: a consonant cluster requires greater effort than a consonant-vowel sequence and may be reduced or replaced at the expense of maximum communication; in addition, co-articulation by near articulators is disfavoured; phonemes of constriction (consonants) give clearer communicative distinctions than phonemes of aperture (vowels) – that is why there are more consonants than vowels in language – but they require more articulatory control (hence the ideal CV syllable).

(4) Reduplication (chronology 2:0 → 2.5):
   phoneme: kadúr → kúdu ‘ball’  dúbi → búbí ‘teddy-bear’
   syllable: sáfta → tátta ‘grandmother’  ejnát → nána ‘Eynat’ (girl’s name)

Explanation: the reduplication often comes as a means to avoid more difficult sound combinations and/or to maintain the number of syllables in the word; sequences of phonemes with the same articulators are disfavoured unless their juxtaposition is, by virtue of some other factor, mutually beneficial. We also found that newly acquired sounds were often reduplicated as a means of practice or of hypercorrection in the clinical situation.

(5) Epenthesis: addition of segments (usually an unstressed) vowel:
   vowel: maftéax → mapatéa ‘key’ (+ stopping, + final consonant deletion)
   (child with CP)
   consonant: toxná → toxsná, toxnít → toxsnít ‘program’ (television/computer)
   (two adult aphasics with Broca’s/Conduction Aphasia)
Explanation: The additional vowel/consonant often eases the transition to more difficult consonants or clusters. The clusters then may be reduced at the expense of communication. Epenthesis as well as the following Assimilation Processes in 7.2(1) below, may also be explained by the factor: sequences of phonemes with the same or near articulators are disfavoured unless their juxtaposition is, by virtue of some other factor, mutually beneficial.

7.2. Assimilation processes (consonant/consonant–vowel harmony)

(1) Velar/nasal/labial, etc. assimilation (chronology 2:0 → 2:8):
 béged → géged ‘clothing’
dubón → dubó ‘little bear’
mángo → bámo ‘mango’

Explanation: a non-velar/nasal/labial sound changes to a velar/nasal/labial because of the influence of or the domination of a velar/nasal/labial sound which entails fewer articularatory gestures at the expense of maximum communication.

(2) Devoicing of final consonants (chronology 2:0 → 3:1):
word: bérez → béres ‘faucet’
jixtóv → ítóf
(+=initial and medial consonant deletion) (child with Dyspraxia)
syllable: taví → tafí ‘ bring’
kóva → kófa ‘hat’ (child with Dyspraxia)

Explanation: additional articulators are disfavoured; voiced consonants become unvoiced in word-final (or medial) position: where the communicative force is less important or crucial, the speaker opts to activate one set of articulators rather than two. This may also be related to the phonetic fact that vowels are shorter before voiceless rather than voiced consonants.

7.3. Substitution processes

7.3.1. Processes reflecting the substitution of active articulators

(1) Fronting: back (non-apical) consonants are substituted by apical consonants usually preserving the same manner and voicing values (apicalisation), e.g. k → t, g → d, f/x → s(θ) (chronology 2:0 → 3:5):
k/g → t/d: kóva → tóva ‘hat’, dag → dad ‘fish’
f/x → s(θ): falóm → salóm ‘shalom’, xátul → satúl ‘cat’, fev → ðev ‘sit’

Explanation: the apex is the most flexible and easy to control of all the active articulators: the earliest and most frequent examples of the substitution of active articulators are
fronting or apicalisation which sharply reduces the number of communicative distinctions of the speaker.

(2) Backing: the back pronunciation of front sounds (usually consonants):

- **Consonants:**
  - mapít → maθít ‘napkin’
  - kos → koʃ ‘cup’
  - batsék → axé (+ initial/final consonant deletion)
  (child with Dyspraxia)
- **Vowels:**
  - éjfo → ófo ‘where’
  - lexem → laʃem ‘bread’
  (child with Down’s Syndrome)
  - aní → aná ‘I’
  (child with Cerebral Palsy)

Explanation: a (possibly idiosyncratic) later, less frequent process where the dorsum (or other back articulators) replaces the apex (or other front articulators; often found in children who have difficulty controlling the musculature of the apex, for organic or other reasons, and/or try to reduce the number of communicative distinctions made by the apex, or other front articulators, especially as a result of earlier fronting or other processes).

7.3.2. Processes reflecting the substitution of turbulence and/or airflow

(1) Stopping: fricatives/affricates are replaced by stops: variable chronology depending on sounds and language (chronology 2:0 → 5:0+):

- sus → tut ‘horse’
- xatulá → katulá
- tsav → tav ‘turtle’

Explanation: maximum constriction is favoured particularly when mobile (stop) phonemes of constriction are easier to control than stabile (fricative) phonemes of less constriction which require greater control of the musculature to create and maintain a small aperture for a stronger turbulent airflow: the most frequent manner substitution for children.

(2) Gliding of liquids: l/r → j/w (variable chronology 2:0 → 5:0+):

- delet → dejet ‘door’
- haláx → hajáx ‘he went’
- mazlég → mazwég ‘fork’

Explanation: maximum aperture is favoured: substitution of a lower to a higher degree of aperture (from consonants to semi-vowels) which may also require less articulatory control.

(3) Glottal replacement: oral consonants are replaced by glottal stop:

- varód → vaʔód ‘pink’
- efrónót → efʔonót ‘pencils’
- agalá → agaʔá ‘wagon’
- ve-lo → ve-ʔo ‘and-no’
- marák → ʔaʔák ‘soup’
  (child with Down’s Syndrome)
Conclusion: additional articulators are disfavoured; a glottal stop replaces an intervocalic consonant or a consonant in syllable-final position; articulatory control of one set of articulators rather than two in an appropriate phonetic environment.

8. Quantitative results and principles obtained from PHB

All of the research performed in the PHB framework yielded quantitative results, which support the following principles concerning the non-random distribution of phonemes in and across languages (and the natural processes which may be associated with them):

(1) additional articulators are disfavoured (devoicing, denasalisation, glottal replacement);

(2) co-articulation by near articulators is disfavoured (consonant cluster reduction, deletions, substitutions);

(3) co-articulation by the same articulators/phoneme is even more highly disfavoured (unless their juxtaposition is, by virtue of some other factor, mutually beneficial – see point 13 below) (reduplication, assimilatory processes, deletions, substitutions);

(4) different word (stem or root) positions have different communicative force and thus affect the favouring and disfavouring of different articulatory gestures, and the number and type of acoustic features and the distribution of phonemes (final consonant deletion, devoicing of final consonants, deletion of unstressed initial syllables, initial epenthesis);

(5) apical articulations are favoured in general and in final position in particular (fronting, apicalisation);

(6) visual articulations are favoured (particularly in word/stem/root initial position (labialisation);

(7) explosive (mobile/stop) phonemes are favoured in initial position (affrication);

(8) turbulent (stable/fricative) phonemes are favoured in final position (spirantisation, deletions, substitutions);

(9) transitions from one distinct constriction to another within a single phoneme (affricates) are disfavoured (deaffrication);
consonant clusters are restricted concerning different articulatory and acoustic features (e.g., mobility/stability) (cluster reduction, stricture, aperture, airflow substitutions, and deletions);

among constrictions, maximal constriction is favoured (stopping);

among apertures, maximal aperture is favoured (vocalisation, gliding);

sequences of phonemes with the same articulators are disfavoured unless their juxtaposition is, by virtue of some other factor, mutually beneficial (reduplication, epenthesis, assimilation processes, substitutions).

The following principles were obtained from the developmental and clinical research performed within the theory of PHB:

the preservation of as many distinctive features as possible (usually 2 out of 3) in substitution processes which require more effort than deletion processes;

the preservation of as many communicative oppositions as possible in the original word (e.g. the number of phonemes per word) in substitution processes which require more effort than deletion processes;

the use of a readily available phoneme in the speaker’s repertoire in accordance with the immediate phonetic environment in substitution processes which require more effort than deletion processes;

the preservation of the original phonetic structure of the word in deletion processes not involving syllable reduction and in reduplication;

if the original structure of the word is reduced by the deletion of syllables, the stressed syllable bearing the most communicative information is maintained;

if the original structure of the word is enlarged by epenthesis, the epenthesis makes the transition to or between more difficult sounds easier.

These principles empirically support the following conclusion underlying the theory: Developmental and clinical speech errors may be viewed as an extreme version of the mini-max struggle: there is less than maximum communication because of either extreme minimal effort or a lack of control over the articulatory tract or mechanisms. Greater effort will be exerted in order to achieve more efficient or better communication through clinical intervention.
8. Summary and conclusions

To summarise the major principles of the theory of PHB for phonological analysis (and for developmental and clinical phonology):

(1) We begin with the phonetic observations, articulatory and acoustic, within which there are no observable units (child language inventory and clinical intake).

(2) By means of the communication orientation we can establish the number of distinctive units of a language (as found in a child or client).

(3) Consideration of the acoustic and physiological characteristics of the units (and the ones found in a child and client) suggests a variety of characterisations.

(4) In choosing among these characterisations, it is apparent that the characteristics of the units must be of such a kind that the human user can learn them (both in typical developmental and atypical language acquisition).

(5) We do not know in advance, deductively, in exactly what way the human factor will interact with the communication and other factors.

(6) Phonotactic skewings in language (diachronic and synchronic) (and in typical and atypical developmental data) reflect the learning process of the speakers (including children and clients).

(7) This skewing viewed consistently with the human factor against the background of the other factors of communication, acoustics and physiology informs us of the characteristics we are confronted by.

(8) Phonology is not random but motivated; the frequencies of the phonological units and the ways they combine are determined both by their phonetic make-up and by the speaker’s (child’s or client’s) exploitation of – or coping with – that make-up in the act of communication and/or by the demands posed by the listener for the message to be understood.

(9) Gestures enhancing communicative distinctiveness are favoured and articulatorily more difficult gestures are disfavoured.

(10) There is a conflict between the communication and the human factors in the striving for maximum communication with minimal effort in the diachronic development of a language and its current synchronic state.
This conflict is even more keenly felt in language acquisition where functional errors and processes may be observed and even more so in the clinic where typical and atypical developmental errors and processes show an even more extreme conflict between the communication and human factors.

The theory of PHB can explain the connection and interrelationship between the phylogeny, ontogeny and pathology of the development of sound systems in human language in a principled way. (One must also take into account that there are some speech disorders, such as dyspraxia or ataxia of speech, where the articulatory errors are generally not considered to be predictable/systematic, whereas for other speech disorders, the errors are recognised as being more systematic and more “lawful”. It must also be noted that a speech disorder may change dramatically as a function of the task being performed, the severity of the disorder, the nature of the disorder as well as the social context of the speech act.)

REFERENCES


Adam, G. 2002. From variable to optimal grammar: Evidence from language acquisition and language change. (Unpublished PhD dissertation, Tel-Aviv University.)


Buk, E. 2003. Phonology as human behaviour: Inflectional morphology in Old Church Slavonic, old and modern Russian. (Unpublished MA thesis, Ben-Gurion University of the Negev, Be’er Sheva.)


Davies, J. 2006. “Phonology without the phoneme”. In: Davis, J., R.J. Gorup and N. Stern (eds.). 163–175.


Tubul, R. 2002. Inflectional systems in Ladino according to the theory of phonology as human behaviour. (Ms., Ben-Gurion University of the Negev, Be’er Sheva.)

Trubetzkoy, N. 1939. Grundzüge der Phonologie. (Travaux du Cercle Linguistique de Prague 7.)


Address correspondence to:
Yishai Tobin
Department of Foreign Literatures and Linguistics
Ben-Gurion University of the Negev
P.O. Box 653
Be’er Sheva 84105
Israel
yishai@bgu.ac.il