From Sound to Meaning in Context

Studies in Honour of Piotr Ruszkiewicz

Edited by Alicja Witalisz

8

Text – Meaning – Context:
Cracow Studies in English Language, Literature and Culture

Edited by
Elżbieta Chrzanowska-Kluczewska
Władysław Witalisz

PETER LANG EDITION

The Relationship between Phonology and Phonetics

Eugeniusz Cyran
John Paul II Catholic University of Lublin

Należy fonetykę od fonologii odróżniać, ale nie należy ich oddzielać.
Phonetics and phonology should be told apart, but not taken apart [EC].

(Stieber 1955: 73)

1. Introduction

In his brief note, Stieber lays out his views on the relationship between the two components of language by saying that "phonological considerations which are not based on phonetic studies hover in the air [...] on the other hand, phonetic studies which do not aim at a phonological synthesis are practically pointless [EC]".[1] At first, the above quotations look paradoxical. How can two things be told apart if they cannot be taken apart, and studied separately? What are the criteria for deciding that a given phenomenon is phonological or phonetic? What is the nature of their relationship if phonology and phonetics are indeed autonomous?

Any discussion of the relationship between phonology and phonetics assumes implicitly that these are indeed two separate entities. Stieber warns us, however, against two opposite perspectives in the way phonological or phonetic studies can be carried out, which he considers bad practice. Phonologists are advised to ground their work in phonetics, while phoneticians are cautioned that their endeavours should serve a higher purpose – phonological synthesis, whatever that is. More than half a century later, the above questions are still relevant and the debate concerning the relation between phonology and phonetics is even further away from a solution. A variety of points of view exist, including those which exclude one of the two components from grammar. Consequently, the very question of the relationship between phonology and phonetics becomes immaterial. Thus, on the one hand, there is the strong position of Ohala (1990), who maintains that there is no interface between phonology and phonetics because phonetic theory itself is sufficient to deal with the observed sound patterns in languages. Though largely correct – phonetic theory is indeed making constant progress and is outstripping phonology in more and more areas to do with the organization and behaviour of speech sounds – Ohala still seems to distinguish the two fields, if only

1 "Rozważania fonologiczne nie oparte na badaniach fonetycznych wiszą w powietrzu [...] zaś badania fonetyczne nie dążące do fonologicznej syntezy są właściwie bezcelowe" (Stieber 1955: 73). It was Professor Piotr Ruszkiewicz who drew my attention to this quote.

terminologically, by saying, for example, that "[...] phonetics offers one of the most obvious paths between phonology and other disciplines" (Ohala 1990: 165). What is phonology then? It appears that by denying the existence of an interface between phonology and phonetics, and advocating a close integration between the two fields, Ohala is no less paradoxical than Stieber. Or, to put it differently, he might be talking about the same thing as Stieber, though from a strictly phonetic perspective which is covering more and more ground in the 'universe of speech', allowing for phonological synthesis, yet not offering a sharp definition of what phonology is, or should be. Ohala, then, responds positively to the postulate that phonology and phonetics should not be taken apart, and offers only a phonetic perspective on how to tell them apart.

On the other extreme, we find proposals such as Hale and Reiss (2000, 2008) who draw a sharp line between form and substance, excluding phonetic substance from phonology, and arguing that the latter is a computational module of grammar, while the former is not.[2] Their position is sharply and eloquently laid out in the following quotes, of which the first one seems to be compatible with the views of Stieber. The second one, however, suggests that phonology and phonetics should be taken apart, or does it?

> The modular approach to linguistics, and to science in general, requires that we both model the interactions between related domains and sharply delineate one domain from the other (Hale – Reiss 2000: 158).

> Phonology is not and should not be grounded in phonetics since the facts that phonetic grounding is meant to explain can be derived without reference to phonology. Duplication of the principles of acoustics and acquisition inside the grammar violates Occam's razor and thus must be avoided. Only in this way will we be able to correctly characterize the universal aspects of phonological computation (Ibid.: 162).

On a closer inspection, what Hale and Reiss say is not incompatible with either Ohala's or Stieber's views. It is simply a different, phonological, perspective, one that does not ignore the results of phonetic research. On the contrary, it seems to embrace it happily because the core of the substance-free research programme in phonology is that substance-based speech sound patterns should have a phonetic explanation only. Consequently, pure phonology is a computational module which is much smaller than it is generally assumed, but it does exist as separate from phonetics. An additional and long-standing argument in favour of substance-free phonology mentioned by Hale and Reiss is based on the fact that phonology must be modality-free as there is such a thing as the phonology of signed language. All this, however, does not mean that some way of relating phonetics with phonology,

2 For a recent survey of a broader range of proposals see, e.g., Kingston (2007).

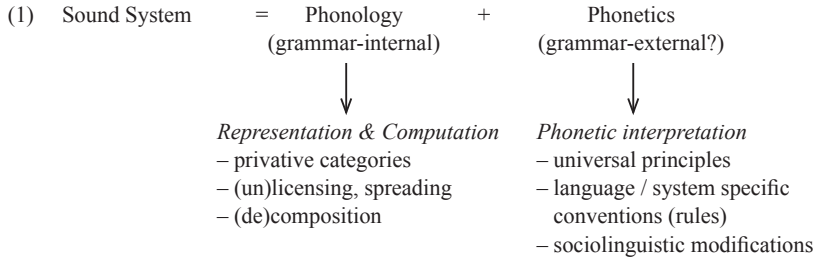
just as signs and phonology, should not be sought. Whether this type of phonological practice 'hovers in the air' is then an empirical question.

In this paper, I attempt to fully embrace the spirit of the views of both Stieber and Ohala by working from a phonological perspective similar to Hale and Reiss's. The seeming paradox in the first quote in this paper calls for a no less paradoxical solution. I start with the assumption that phonology and phonetics cannot be told apart (delineated) if they are not taken apart first. How they interact is another issue, which will also be addressed. Firstly, I begin from a phonological perspective of Government Phonology (GP), which seems to be a good candidate for a substance-free model, if some modifications are implemented. Secondly, I assume phonology and phonetics to be autonomous (told apart), yet interacting in a conventionalized way to form a sound system (not taken apart). It will be argued that most of the confusion in the discussion on phonology and phonetics stems from the fact that sound systems are mistaken for phonology.

2. Sound system, phonology and phonetics

In the 'universe of speech', a sound system is the sum total of phonological and phonetic aspects which together are responsible for the observed phonetic facts. In this model, a sound system cannot be identified with phonology, because that would ignore phonetics. Neither can a sound system be identified with phonetics only. In other words, a sound system stands behind the observed phonetic facts in a given language, but it cannot be directly identified with phonetics. As a consequence, phonetically observed facts are not entirely independent of the particular system in which they occur. Phonetic facts are always a result of phonetic interpretation of phonological representation. They follow from the system, and as such they may be ambiguous and misleading. To understand a sound system, one has to find out how phonology and phonetics interact in that system. Both phonology and phonetics are separate and can be studied separately, but when sound patterns or systems are taken into account, the two aspects must dove-tail to produce the results. The graph in (1) and the discussion below further clarify how phonology relates to phonetics in a sound system.

(1)



2.1. Phonology

The phonological side of the equation comprises representation and computation, that is, a phonological structure organizing a set of symbols, and principles of their manipulation. For reasons of space and relevance, the discussion of representation is restricted to melodic primes (elements), while prosodic structure is left out. A concrete illustration of the proposal will be based on the laryngeal system(s) of Polish.[3]

The privativity of phonological categories which is assumed here has been argued for elsewhere and does not require additional argumentation (see e.g. Avery 1996; Harris 1994, 2009; Honeybone 2002, 2005; Iverson – Salmons 1995; Lombardi 1991, 1995). Likewise, not much needs to be said about the computation, especially within GP. In this theoretical model, segments are composed of one or more elements and require licensing. Under insufficient licensing conditions, for example, due to a particular prosodic context, segments may be decomplexified (decomposition), while processes of spreading of categories may lead to addition of elements to existing representations of segments (composition). Below, I provide a simplified and rather uncontroversial illustration of the four instances of processing operations: licensing, decomposition, spreading, and composition, ignoring details which are irrelevant to the discussion. (2a) shows voice assimilation as composition due to element spreading. I follow, e.g., Gussmann (2007) in assuming that it is the laryngeal element {L} that is responsible for the voice contrast among Polish obstruents. The voiceless series is unmarked. In (2b), we observe a phenomenon of final obstruent devoicing (FOD) in Polish as decomposition due to weak licensing in the word-final context.

(2) a. voice assimilation

p	r	ɔ	ɕ	+	b	a
			|		|	
			<	<<<	L	

prośba [prɔʑba] ‘request’

b. FOD *weak licensing*

m	a	ʑ	#
		ǂ	
		L	

maź [maɕ] ‘sticky substance’

Given that the voicing contrast in Polish is indeed expressed by the presence of a privative element {L} in the representation of voiced obstruents, the processes illustrated above can be described in the following way. If /ɕ/ is composed of elements {x,y,z}, its voiced congener /ʑ/ is one element more complex, that is, {x,y,z,L}. In [prɔʑba], the laryngeal element is spread from the following ob-

3 One should probably use the term ‘laryngeal sub-system’ here, equating ‘system’ with language and allowing for a number of such sub-systems to be part of a larger system involving various dimensions, for example, vocalic, place, manner, laryngeal, etc.

struent. On the other hand, in [maɕ], {L} was present lexically, but delinked. FOD is a case of decomplexification under weak licensing and turns {x,y,x,L} into {x,y,z}. This is more or less the essence of privative analyses of such phenomena. It should be emphasized that the unmarked (voiceless) obstruents do not receive any further specification – they are interpreted as voiceless unaspirated if {L} is not present in the representation.

What is more important for our discussion is how the categories receive their phonetic definition in a substance-free phonology. The answer to this question will not change much of the above analysis because we can always assume that the set of symbols we use to discuss phonological phenomena, willy-nilly, must already contain information as to what a given phonological category corresponds to in the real world (of phonetics). Nevertheless, a possible way of looking for an answer will be offered below.

The question of substance acquisition relates to one of the three main points of interaction between phonology and phonetics (Kingston 2007). In discussions of the definition of distinctive features, the typical problem is whether they are articulatory, acoustic, or auditory, or in fact, whether they could holistically involve all types. An imminent verdict on this issue is unlikely, and, as I will argue below, unnecessary. From our perspective of the relationship between phonology and phonetics a more important question seems to be whether phonetic theory can model the emergence of the substance of the distinctive features. Whether melodic primes are emergent and need not be postulated to be innate is not a problem for substance-free phonology. The question that remains then is: what is a feature, a categorical distinction, without substance? Our tacit assumption at this stage will be that it is simply a decision to use an additional contrastive dimension by assigning a new privative category to one of the resultant contrastive series. The property will be given flesh by a systemic interface with phonetics. For example, in the case of the /ɕ–ʑ/ contrast, where {x,y,z} constitute the common denominator, it is a matter of introducing a fourth element, or a fourth dimension of contrasts.[4]

One of the functions of phonology is to define categorical contrast. In privative models this boils down to a presence or absence of a particular property to distinguish two segments. If no contrasts are used in a particular dimension, e.g. laryngeal, then one series of obstruents is typically found – the voiceless unaspirated, e.g. Hawaiian.[5] We may assume that such languages do not use any laryngeal elements, a fact that will be represented below with a superscripted zero next to C, which stands for an obstruent, that is, (C^{o}).

4 It need not be stressed that {x,y,z} are not real elements or dimensions. What exactly makes up the fricatives is not relevant here.

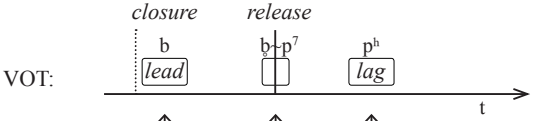
5 In considerably fewer cases, it is the voiced series, e.g. in Yidiny (Keating – Linker – Huffman 1983).

In languages like Polish, or Icelandic, which have a two-way laryngeal contrast among obstruents, that is, between a voiceless unaspirated and fully voiced one for Polish, and between a voiceless aspirated and voiceless unaspirated one for Icelandic, only one laryngeal element is used. In Polish, the marked representation involves the presence of {L} in the voiced series (Gussmann 2007), while the neutral series is 'toneless' (C^{L} vs. C^{o}). In Icelandic, on the other hand, the distinction is that of (C^{H}) for the aspirated series, and (C^{o}), again, for the voiceless unaspirated one. This, in essence, is the Laryngeal Realism view (Honeybone 2002, 2005; Harris 1994, 2009; Gussmann 2007). For completeness, one may add two other types of systems: one with three and one with four contrastive series, which can also be represented only with the two laryngeal elements mentioned above. Thai contrasts three series /b,p,p^{h}/, while Hindi has a four-way contrast /b,p,p^{h},$b^{ɦ}$/.

The privative representation of laryngeal contrasts in GP, which uses only two elements for this dimension, including the Hindi case, appears to be compatible with the finding of Lisker and Abramson (1964) that there are three major phonetic categories which are utilized by languages, and which, as in the phonetic descriptions above, can be quite elegantly illustrated by means of points or regions along the VOT continuum. Firstly, there is full voicing, which can be referred to as long lead or negative VOT. This property corresponds directly to the phonological element {L} in Laryngeal Realism. Secondly, there are consonants characterized by a short lag (voiceless unaspirated stops). This phonetic realization seems to correspond most readily to the neutral obstruent (C^{o}) which is present in all four systems mentioned above. Finally, consonants may have a long lag (voiceless aspirated stops). The element responsible for this distinction is {H}.

The relationships between the phonetic contrasts most commonly used in languages and the GP elements is illustrated by the following graph.[6]

(3)



	C^{L}	C^{o}	C^{H}	
Hawaiian	–	C^{o}	–	
Polish	C^{L}	C^{o}	–	
Icelandic	–	C^{o}	C^{H}	
Thai	C^{L}	C^{o}	C^{H}	
Hindi	C^{L}	C^{o}	C^{H}	/ $b^{ɦ}$/ = C^{L+H}

6 The common practice is to use plosives in illustrations of VOT.

7 In aspiration languages, spontaneous voicing (also called passive voicing) may occur. The term 'passive voicing' will be explained further below.

Each system has to have the unmarked series /C°/. It should be noted, however, that this involves a range of realizations from slightly voiced to voiceless unaspirated. A laryngeal element, either {L} or {H} appears only if a two-way contrast needs to be represented. Thai and Hindi utilize both elements, but the latter language allows them both to be present in a single segment. [bʱ] is a plosive which begins with a long lead and ends with a long lag. The English system, for comparison, is phonologically similar to Icelandic in that it is assumed to use {H}. However, its neutral obstruents are often realized with some voicing, also called passive voicing. It is interesting to note, anticipating a little the discussion of phonetic interpretation, that the passive voicing is possible only in 'aspiration' languages using {H} and impossible in L-systems. There seems to be an asymmetry between voicing and aspiration languages (Lisker – Abramson 1964), in that fully voiced obstruents do not contrast with partially voiced ones, while voiceless unaspirated can contrast with voiceless aspirated (e.g. Icelandic). Thus, one contrastive region can be established on the VOT lead side, and two on the lag side. This asymmetry may follow from the general phonetic fact that both perceptually and articulatorily it is difficult to contrast fully voiced with slightly voiced objects, and to control degrees of voicing.

So far, we see an almost biunique relation between the three phonetically defined contrastive values along the VOT continuum corresponding directly to three possible representations of stops in element theory, where full voicing in the signal corresponds to the presence of {L}, aspiration relates to {H}, and voiceless unaspirated objects are typically neutral. Thus, given the phonological marking that is used in a given system it is directly obvious what phonetic values will be used to express it, and vice versa: long VOT lead in the signal suggests the presence of {L} in the representation of a given obstruent, while aspiration leads us into thinking that it is connected with {H}. If the Laryngeal Realism view were correct, phonological representation would always be unambiguously read off from the spectrogram, and phonetic interpretation would be rather trivial. However, the main problem with this model is that it does not work, at least for one of the two major dialects of Polish, as will be shown below. First, let us look at the phonetic side of the sound system.

2.2. Phonetics and phonetic interpretation

The phonetic side of the sound system presented in (1) contains principles of a varying degree of generality rather than importance. Their role is strictly related to phonetic interpretation of phonological representation. First, a distinction needs to be drawn between universal phonetic principles and universal principles of phonetic interpretation. These terms are not synonymous. The former relates to

physiology of speech and to phenomena which can be studied independently of phonology. One example of such a principle is the general aerodynamics leading to spontaneous vibration of vocal folds and its inhibition. The second term – universal principles of phonetic interpretation – is ambiguous and misleading. It suggests that phonological representation[8] contains universal instructions as to how it should be pronounced. This would be compatible with the Laryngeal Realism view presented above. However, it seems that the direction of motivation may be the reverse. Phonetics provides options of phonetic interpretation, which are selected or associated with particular phonological categories in a chiefly arbitrary fashion. Secondly, phonetic interpretation is always system dependent, that is, language specific, rather than universal. Nevertheless, if we understand phonetic interpretation as a relation established between phonologically defined categorical contrasts and the phonetic contrastive regions, as in the case of the three regions along the VOT continuum, we could also identify what appears to be a universal principle of phonetic interpretation: the principle of sufficient discriminability in production and perception. It is universal in the sense that most known languages seem to follow it.[9] The universality in terms of production is guaranteed among humans due to physiology. On the other hand, the same cannot be said about perception. It may be possible to phonetically define a universally potential maximal number of phonetic contrasts in a particular dimension. Just as in the case of the VOT continuum, it is possible to define such contrasts in the vowel space as well. The actual perception of speakers is always curtailed by the particular system they have acquired. Thus, perception is to a great extent language specific, unless we want to talk about potential and not the actual sound systems.

Returning to the VOT contrasts, phonetics provides regions which allow for minimal phonetic distance and therefore for discrimination. However, it is the role of phonetic interpretation conventions to express the categorical distinctions provided by phonology. This is where the universal principle of sufficient discriminability, or better, sufficient *phonetic distance* comes into play, which is to a great extent dependent on the number of contrasts demanding expression in a given phonetic space (see, e.g., Liljencrants – Lindblom 1972). Thus, for example, languages with a two-way laryngeal contrast tend not to select the maximally dispersed phonetic categories: long VOT lead (fully voiced) contrasts with short lag (voiceless unaspirated), rather than with long lag (voiceless aspirated).

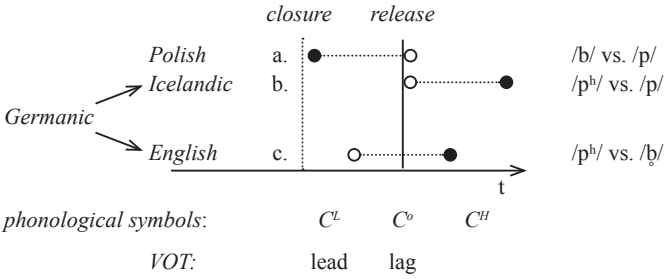
---

8 It is possible that we also need a distinction between phonology proper, as a substance-free computational module, and phonological representation (including lexical representation), in which substance in the sense of established connections between subsegmental representation and phonetic interpretation are present.

9 The principle has been applied, for example, to the understanding of vowel systems (see, e.g., Schwartz – Boë – Vallée – Abry 1997).

Sufficient phonetic distance does not only mean that contrasts need not be maximized, it also means that there is something like a minimal distance. One example of this has already been mentioned with respect to the VOT continuum. Namely, no contrasts between full voicing and partial voicing (long and short VOT lead) can be found (*/b–b̥/). Another interesting example concerns the interpretation of the neutral obstruents (C^{o}) in English and Icelandic in relation to the marked congener (C^{H}). Recall that the neutral obstruents may be passively voiced in English and tend not to be so in Icelandic. This fact coincides with the phonological and phonetic robustness of aspiration in the two languages. Icelandic aspiration is stronger than in English, both acoustically and in terms of phonological behaviour. It tends to survive in more contexts than in English, and may be subject to temporal shifts rather than loss, a phenomenon called pre-aspiration (Gussmann 1999). The observation which is relevant to this discussion is that robust aspiration minimizes the chances for passive voicing, and vice versa. It appears then, that the relation between the two obstruent series in these languages observes something like a sufficient distance, where both the marked and the unmarked series are subject to a coordinated variation. The following graph attempts to express the main points of our discussion so far. Below, the black circle denotes the marked obstruent series, while the white circle corresponds to the unmarked congener. The dotted line between the black and white circles indicates the sufficient phonetic distance, which is rather symbolic, and it cannot really be measured in, e.g., temporal units as the graph may suggest. The slight shift of the English marked-unmarked pair to the left in comparison to Icelandic indicates that the aspiration is less robust and that the unmarked series may be subject to passive voicing. It is passive; it will be recalled, because there is no phonological category standing behind it. It is merely a systemic interpretational phenomenon.

(4) Phonetic distance and variation



Universal and language specific principles of phonetic interpretation do not seem to have a clear boundary, especially with respect to certain types of segments,

for example, obstruents. Let us dwell a little on the question of the aerodynamic conditions inducing vocal fold vibration (voicing) in order to be able to show how this universal phonetic principle is affected by systemic (language specific) considerations.

The vibration of vocal folds occurs under special aerodynamic conditions involving a number of articulatory parameters. The desired effect of these articulatory settings is to achieve a sufficient drop in air pressure and air flow between trachea and pharynx (Chomsky – Halle 1968; Halle – Stevens 1971). The drop in air pressure is inhibited in segments which are produced with some narrowing in the vocal tract because occlusion leads to intra-oral air pressure build-up. These simple physical facts are responsible for the so called universal markedness tendency for vowels and sonorant consonants to be voiced and for obstruents to be voiceless. In phonological descriptions, these simple aerodynamic facts are often expressed by the use of the following default rules (e.g., Gussmann 1992: 43; Rubach 1996: 77, 80).

(5) a. [sonorant] → [+voice]
b. [obstruent] → [–voice]

While (5a) seems to be overwhelmingly correct – vowels and sonorant consonants are typically voiced, obstruents seem to defy the supposedly phonetically natural rule in (5b). First of all, sonorant voicing is generally considered spontaneous, which has led to proposals that it should not be expressed phonologically by means of any feature or element. This is also the position of a number of privative feature frameworks, including the Laryngeal Realism view and the model presented here.

As for obstruents, under certain articulatory and contextual conditions they also may be spontaneously voiced (Westbury – Keating 1986). We may generally describe these conditions as lenis articulation[10] and voiced environment, that is, adjacent vowels or sonorant consonants. Due to the fact that, unlike in sonorants, such voicing is dependent on the environment, instead of spontaneous, the term passive voicing is often used to refer to this situation (e.g. Kohler 1984). Westbury and Keating (1986) note an interesting paradox about some of the languages possessing only one series of obstruents. Recall, that this concerns the segments which we symbolize as C^{o}, that is, laryngeally unspecified ones, which are typically realized as voiceless unaspirated. This voiceless articulation is maintained also in contexts (voiced environment) in which spontaneous voicing would be

10 These include, for example, relatively short closure, contracting the respiratory muscles, decreasing the average area of the glottis and/or tension of the vocal folds, decreasing the level of activity in muscles which underlie the walls of the supraglottal cavity, actively enlarging the volume of that cavity, etc.

phonetically more natural. It would seem then, that the default rule (5b) above may in some cases be phonetically unnatural. Westbury and Keating acknowledge that this lack of voicing is due to 'more powerful principles', for example, a systemic tendency to maintain the phonetic similarity among the positional allophones. Clearly, these more powerful principles override natural phonetics and should be viewed as stemming from the interaction between phonetics and phonology. These are phonologically dependent decisions on the phonetic interpretation of segments.

To conclude this part of our discussion. Thus far, it is clear that phonetic theory alone turns out to be insufficient in the study of sound systems. Likewise, without phonetics providing predictable contrasting regions, substance-free phonology would be equally lacking. It appears that phonetic interpretation is not a phonological instruction. Rather, it is an interface phenomenon. Phonology provides the number of contrasts, while phonetics provides the possible phonetic contrasts. The relation between the two could potentially be quite arbitrary,[11] though complying to some principles. Below, I relate another argument in favour of the arbitrariness of the relation between phonology and phonetics in sound systems.

3. Sandhi voice assimilation in Cracow-Poznań Polish

Polish divides into two dialect groups with respect to voicing: the Cracow-Poznań (CP) and Warsaw (WP). The phonetic and phonological facts in these dialects seem to be generally identical, except for the assimilation phenomena across word boundary, the so-called sandhi voicing. Thus, both dialects have a two-way laryngeal contrast between obstruents involving fully voiced and voiceless unaspirated congeners. Both dialects boast the same processes, such as final obstruent devoicing (FOD) and word-internal voice assimilation (VA). However, the voice assimilation phenomena in the external sandhi context are markedly different. In CP, a word-final obstruent becomes voiced before any voiced segment beginning the following word, that is, before a voiced obstruent, a vowel, or a sonorant consonant. In WP, on the other hand, voice assimilation occurs only if the following word begins with a voiced obstruent. The facts are independent of the lexical representation of the final obstruent.

The data below show lexically voiced and voiceless stops in the context before another word beginning with a vowel (_V^{+v}), a sonorant consonant (_S^{+v}), a

11 The arbitrariness of these relations deserves a longer discussion which will have to be reserved for another occasion. Suffice it to say that, like in the acquisition of vocabulary, the relation between the concept and the phonological form that expresses it is not at all arbitrary to the learners of their language. It is rather arbitrary for the linguist.