DISCOVERERS OF THE PHONEME

HARRY VAN DER HULST

8.1 OVERVIEW[1]

THE central insight that underlies the linguistic study of speech is that sounds which are objectively different can count as the same at some level of linguistic analysis (and most likely also in the mind of language users). The sameness is captured by the notion 'phoneme', an abstract notion that is manifested in a large (perhaps endless) variety of speech sounds. We usually associate the first explicit statements of this insight, which I will call 'the phonemic principle,' with the birth of phonology, but the idea as such is much more widespread and older. While earlier writers and traditions will be mentioned in the introductory §8.2, §8.3 will focus on developments that began toward the end of the nineteenth century with Jan Baudouin de Courtenay and Mikołai Kruszewski (the Kazan School), and the impact of Ferdinand de Saussure's structuralist ideas, which together culminate in the Prague School (with N. S. Trubetzkoy and Roman Jakobson as the leading figures), discussed in §8.4. §§8.5–8.7 discuss Louis Hjemslev's glossematic theory, the London School (the prosodic analysis of J. R. Firth), and various American structuralists (Franz Boas, Edward Sapir, Leonard Bloomfield, and the 'post-Bloomfieldians'), respectively. §8.8 reviews later developments in Europe and also covers work in the Soviet Union. §8.9 focuses on Generative Phonology, the theory of Noam Chomsky and Morris Halle, and developments of this approach up to the present time. §8.10 winds up with some conclusions.

Goldsmith and Laks (2000b) make a distinction between an external history (theories, dates, names) and an internal history of a subject; the latter deals with the personal motives, contacts, and general backgrounds that lie behind historical developments.

The present overview is more external than internal since the latter approach requires much more space and most of the research has yet to be done.[2] I have tried to include references to important and pivotal works, both those that are part of the history and those that reflect on it. In dealing with the history of a discipline, or any history for that matter, we cannot always speak of a straight line, and this certainly applies here, since different individuals and schools developed principles of phonological analysis, often with little cross-reference, although it remains to be seen how various scholars influenced each other and to what extent various issues that were 'in the air' developed independently.

My main sources for this overview up to and including the early phase of generative phonology are Fischer-Jørgensen (1975) and Anderson (1985), two excellent overviews which offer detailed discussions of the individual phonological schools. I also mention Kilbury (1976), who presents a thorough discussion of the development of 'morphophonemics'. Many introductions to linguistics (e.g. Dinneen 1967) and general works on the history of linguistics also contain valuable overviews of approaches to phonology (and phonetics). Several other shorter works offer important discussions of the development of phonology, such as Kortlandt (1972), Durand and Laks' (2002) discussion of the cognitive setting of phonology, various papers by John Goldsmith (2000, 2005, 2008), Laks (2001, 2005), (Allan 2010a: ch. 9), and Dresher (2009, 2011). In what follows, I will derive many of my points from (interpretations of) these various works, having made an effort to confirm these remarks from my reading of 'original' works that played a pivotal role in this history.[3] There are various important collections of seminal phonological articles or book chapters (or sometimes excerpts): Makkai (1972) contains many classical papers from the American structuralist school and early generative phonology. See also Joos (1957) for the structuralist period and Vachek (1964) for a reader with Prague School work. Fudge (1973) is a much smaller collection with a broader orientation. Dinnsen (1979) and Goyvaerts (1981) contain overviews of a variety of Generative Phonology that precedes the development of non-linear models for which various volumes edited by van der Hulst and Smith contain representative work (1982ab, 1985a, 1988a, b, c). See also Durand and Laks (1996) for an important collection. Goldsmith (1999) focuses on generative phonology and McCarthy (2004) focuses on Optimality Theory. Kreidler (2000) offers a huge collection of articles covering many different approaches. Very useful also are various handbooks in phonological theory such as Goldsmith (1995), de Lacy (2007), Goldsmith et al. (2011), Kula et al. (2011), and in particular van Oostendorp et al. (2011), which contains 120 long chapters each offering overviews of a different area of phonology. Since 1985 we have

[2] Goldsmith and Laks (2000a) contains a number of studies that focus on internal history, as do several essays by Jakobson (e.g. 1971) and various other reflective works by the key figures in the history of phonology. Fischer-Jørgensen (1975) and Anderson (1985) do much more than enumerate the facts. These works contain numerous insightful remarks and passages on connections, general political and scientific background, 'who talked to whom,' etc.

[3] Some works fulfil both functions in that they review preceding work and then add their own point of view.

had a journal (*Phonology*, from 1984 to 1987 the *Phonology Yearbook*, CUP) devoted to theoretical phonology. Since the early 1990s there have been a number of phonology conferences in Austria of which the proceedings (called *Phonologica* [followed by the year], e.g. Dressler and Pfeiffer 1977) display a broad array of approaches.

Finally, a note on the term 'phonology'. For many, this term stands in contrast to 'phonetics,' and indeed, much of the history of phonology is about its relationship to phonetics. Phonetics, the story goes, studies the physical aspect of speech, i.e. how speech is produced and perceived as well as its physical properties, while phonology studies how speech sounds function distinctively and how these functioning units enter into paradigmatic (systemic) and syntagmatic (sequential) relations. Ignoring for the moment that many phoneticians would argue that their research goes beyond the strictly physical, this leaves us without a term that covers both disciplines. In the view of some (notably the British linguists), phonology is just a branch of phonetics, which others (like André Martinet) designate as 'functional phonetics.' Another view (characteristic of the American schools) is to use phonology as the cover term for 'phonemics' and phonetics. Since, clearly, phonemes (i.e. distinctive units) form the fundamental concept for phonology, I believe that the American practice deserves general use, also because this gives us a term ('phonology') to refer to a continuum of phenomena and research, since the boundary between the 'physical' and the 'functional' is not always so clear. Here, however, in keeping with more common usage, I will use phonology (as phonemics) and distinguish it as such from phonetics.

8.2 INTRODUCTION

The earliest development that is relevant to the history of phonology (that we know about) is perhaps the development of writing, and especially of *phonographic*, more specifically *alphabetic* systems. Then, as today, phonographic writing systems tend to be phonemic in character, avoiding different symbols for linguistically non-functional, allophonic distinctions that are not relevant for differentiating lexical meanings. This is true whether systems are alphabetic or involve larger syllabic units. However, alphabetic systems specifically anticipate the modern view that phoneme-sized units are the pivotal *sequential* components underlying speech. Occasional diacritic characters representing sub-phonemic sound qualities may even be said to anticipate the notion of distinctive features.

Early grammarians in various (Greek, Roman, Indian, Arabic, Chinese)[4] traditions often show explicit awareness of the notion of speech sound (see Allan 2010a: ch. 9 and Chapters 2 and 5 above), and an implicit recognition of the need to abstract away from non-meaning differentiating phonetic properties, but no explicit early developments of

[4] For Greek and Roman see Robins (1953) and Allen (1981); for Indian see Allen (1953), Deshpande (1995), and Ch. 11 below; for Arabic and Hebrew see Semaan (1968); for Chinese see Halliday (1981).

the phonemic principle can be found in these works. Pāṇini's work on Vedic might be an exception (Kiparsky 1979; Chapter 10 below), although closer investigation of other older sources might reveal otherwise. An isolated obvious and well-known exception, discussed by Fischer-Jørgensen (1975: ch. 2) and more extensively in Allan 2010a: ch. 9), concerns the anonymous Icelandic linguist who, in the twelfth century, wrote what is now known as the 'First Grammatical Treatise' in which, in the course of proposing a new writing system, he demonstrated the existence of phonemes (without using this term) by listing sound differences that occurred in otherwise identical environments, thus effectively producing minimal pairs (or rather minimal sets) and using the commutation test which is the fundamental discovery procedure for establishing phonemic contrast. This work, unfortunately, did not attract the attention it deserved.

In the following centuries, to be sure, others addressed the matter of spelling of other languages, also relying on the recognition of *relevant* sound differences. In conjunction with these concerns, these scholars often provided detailed phonetic descriptions of speech sounds (e.g. in the work of Thomas Smith and John Hart in England). Allan (2010a: 199ff.) also mentions various other British scholars who were concerned with spelling and pronunciation. Firth (1946), Abercrombie (1948), Fromkin and Ladefoged (1981), Ohala (2004), and Allan (2010a) also discuss various early (forgotten) phoneticians[5] whose work involves understanding of the distinctive role of speech sounds (often referred to as 'letters') and the fact that these sounds can be analysed in articulatory gestures. Fromkin and Ladefoged refer to William Holder (who offered a detailed description of articulation to be able to teach speech to the deaf), John Wallis, and Francis Lodwick. These scholars developed systems of articulatory building blocks, often with the explicit intention of such inventories being universal. Fromkin and Ladefoged's specific point is that these scholars realized not only that sounds could distinguish meaning but that they perhaps were on the brink of locating this distinctive function with the smaller articulatory ingredients. Ohala (2004) mentions a variety of other phoneticians (Johan Conrad Amman, Wolfgang von Kempelen, Erasmus Darwin, Robert Willis, T. Hewitt Key, and others). Several of these scholars developed experimental methods or built speech-producing contraptions. Interestingly, Erasmus Darwin (as pointed out in Ohala 2004) proposed a system of 13 'unary features' (Darwin 1803).

The nineteenth century shows great developments in the comparative study of languages (usually called 'philology'), with an emphasis on the historical developments of individual sounds or groups of sounds (see Allan 2010a: 207ff.; Waterman 1963, Robins 1967). While such work may generally have not been very precise in making reference to the phonetic properties of these sounds, there are various exceptions, such as Karl Verner (1846–96). It must also be noticed that the numerous studies of the historical development of 'speech sounds' most certainly rested on the *implicit* recognition of phonemic units and

___

[5] See also Ch. 5 above.

distinctive sound properties, even though the distinction between relevant and irrelevant phonetic properties had not yet been made.

Meanwhile, phoneticians developed more sophisticated experimental methods, and toward the end of the nineteenth century these methods allowed them to observe a considerable variability in the realization of speech sounds. As a consequence it no longer seemed justified to speak about speech sounds as the alleged invariant building blocks of words. Clearly, certain differences between speech sounds were more important than others in that some would be negligible because they would not differentiate words, whereas other differences were important in precisely this respect. Once a distinction is made between sound differences that are distinctive and sound differences that are not, we are essentially making a distinction between the phonetic study of sound systems and the phonological (or rather, phonemic!) study of sounds. With this we also have introduced the notions 'phoneme' and 'allophone.' A phoneme is an abstract category that generalizes over a large (infinite) set of actual speech sounds which are its allophones.

It is interesting to note that the founding father of the International Phonetic Alphabet, Paul Passy (1859–1940), felt that an international phonetic alphabet was necessary, precisely to write down speech in terms of symbols that ignore the allophonic differences (Albright 1958, Kemp 1994a). Thus, his 'broad transcription' would essentially be a *phonemic* transcription. A narrow transcription was possible to capture additional phonetic, i.e. allophonic, detail which might be useful (especially as long as the linguist has not yet figured out which properties are distinctive), but runs into the potential problem of not being able to decide where to stop. Another figure who bridges phonetics and phonology was Henry Sweet (1845–1912), who, like Passy, was keenly aware of the fact that one has to recognize that only certain phonetic properties serve the purpose of differentiating between words. Sweet (quoted in Ohala 2004) felt that the new instrumental methods should not take the place of what he called the 'natural method' (based on the perceptual observation of the phonetician), and qualified instrumental phonetics as not being phonetics at all (see Sweet 1877). This makes sense because instruments are by definition not sensitive to a distinction between contrastive and non-contrastive properties while the phonetician (at least in considering his own language) is.

From this point on, phonetics and phonology/phonemics are involved in an ongoing love–hate relationship. Both want to be seen as independent, but neither can do without the other. Ohala (2004) makes a useful distinction between 'taxonomic phonetics,' which culminated in the development of the IPA system, a tool for classifying sounds and their transcription, and 'scientific phonetics,' which aims at understanding the processes underlying speech production and perception. Phonology could easily embrace the results of taxonomic phonetics which, as we have seen, anticipated the phonemic principle. However, phonology did not relate well to scientific phonetics, on the one hand, due to the idea that there is nothing interesting to know beyond identifying the distinctive properties and, on the other hand, because scientific phonetics developed its own set of goals, making use of the rapidly evolving

technological possibilities. One might say that phonology focused on *langue* (or 'competence'), while phonetics, assuming the units of langue/competence, focused on *parole* (or 'performance'), but that would imply a misunderstanding about the appropriate scope of what phoneticians study, many of whom would not insist on this dichotomy to begin with. Ohala (2004) notes that around the 1950s the strict separation, which was rejected even before that or indeed never accepted by several scholars, was bridged by various trends such as the interest in speech synthesis and the rigid development of explicit acoustic and articulatory definitions of distinctive features by Roman Jakobson and others (see §8.4).

8.3 THE FOUNDING FATHERS OF PHONOLOGY

If we see the Prague School views as held by Nikolai Trubetzkoy and Roman Jakobson (and, of course, many other pivotal figures) as a culmination of certain preceding developments, we need to especially discuss two of these which are at least partially independent: the views of the Kazan School (Jan Baudouin de Courtenay and Mikołai Kruszewski) and the views of Ferdinand de Saussure. Fischer-Jørgensen (1975) also mentions other forerunners, such as Otto Jespersen (1860–1943), Adolph Doreen (1854–1925), Johan Forchhammer (1794–1865), and Jakob Winteler (in the late nineteenth century), who had all explicitly recognized that only certain differences between speech sounds were 'functional', being 'linguistically relevant' or 'carrying semantic differences.' This is not to say that the development of phonology was an exclusively European affair. Fischer-Jørgensen (1975) also refers to Edward Sapir as a forerunner who had, in the US, been preceded by Franz Boas (see §8.7).

8.3.1 The Kazan School (Baudouin de Courtenay and Mikołai Kruszewski)

Entirely independently from and prior to de Saussure's work on synchronic linguistics, Baudouin de Courtenay (1845–1929) had embarked on the synchronic study of sound *alternations* in morphologically related words; see Baudouin de Courtenay (1972b), and for a collection of his other work, (1972a). Baudouin de Courtenay worked very closely with Mikołai Kruszewski (1851–1887), an enormously influential student of his.[6] Their work influenced the Leningrad and the Moscow Schools of Linguistics (see §8.8.2) and the founders of the Prague School, but was largely unknown elsewhere in Europe or the

US, although some of Baudouin de Courtenay's students, such as Lev Vladimirovič Ščerba (1880–1944) did come to influence, for example, Daniel Jones in his development of the phoneme concept. However, Baudouin de Courtenay and Kruszewski were aware of de Saussure's earlier work on Indo-European (de Saussure 1879); Baudouin de Courtenay had met de Saussure in 1881 and they subsequently corresponded. De Saussure recognized the resemblances in their work in his notes. Jakobson and other members of the Prague School were also familiar with and clearly influenced by the Kazan School.

Baudouin de Courtenay and Kruszewski recognized the unit phoneme as a generalization over non-distinctive phonetic varieties, but their work focused more on alternations in related words, i.e. what came to be known as 'morphophonology' (in Europe) and 'morphophonemics' (in the US). They acknowledged the distinction between alternations that are purely phonologically/conditioned (both allophonic and neutralizing) and those that rely on non-phonological information, and sometimes use the term 'phoneme' *only* for the units that underlie the former, a usage adopted by the Moscow school. For de Courtenay the phoneme was a psychological unit, and Kruszewski, like de Saussure, saw language as a system of syntagmatic and paradigmatic relations.

Both Kruszewski and de Courtenay developed detailed typologies of different kinds of alternations, essentially boiling down to a three-way distinction into (a) alternations that are governed by fully automatic, transparent, exceptionless, 'phonetic' or low-level rules (allophonic or neutralizing), (b) alternations that follow rules that are no longer fully phonetic but have acquired morphological conditioning (*electric* ∼ *electricity*), and (c) alternations that by themselves encode morphological or semantic distinctions (*was* ∼ *were*). In his detailed discussion of the progression from one type to the other, de Courtenay introduced the term 'phonologization' to refer to the transition of type (a) rules to type (b) rules. Clearly much of the discussion that unfolds in later models, specifically in later developments in generative phonology, was more than anticipated (although unfortunately not often informed) by the work of these two linguists (see §8.9). The Kazan school theories about alternations were known to Jakobson and Trubetzkoy, but, as Anderson (1985: 80) points out, they did not resonate in their work, at least not initially. Rather, the Prague School focused on the phoneme which, although recognized by the Kazan writers, was not their major interest as such, apart from being a relevant unit in the study of alternations.

Several writers (Jakobson 1971 [1960], Anderson (1985: 38, 66–8, Dresher 2011) have pointed out that the term 'phoneme' was invented by the French linguist A. Dufriche-Desgenettes (1804–78), meant as a French equivalent to the German *Sprachlaut*. In his later work de Saussure used the term in this sense, i.e. as standing for 'speech sound', but in his *Mémoire* (de Saussure 1879) it referred to the sound that occurs in the ancestor word for cognates in different languages. From this work Kruszewski adopted the term, but he then extended its use as a reference to pairs of sounds that stand in a synchronic alternation *within* one language. De Courtenay subsequently made a

[6] See Silverman (2012) for a thorough discussion of Kruszewski's views, and Radwańska-Williams (1993).

further step: for him a phoneme was an abstract ('psychophonetic') unit that engages in morphological alternations (allomorphy) and, subsequently, underlies the variable pronunciations of any speech sound, thus also including allophonic variation. De Courtenay did not see his phonemes as built out of smaller units, like features, and therefore did not think of his acoustic images as strictly redundancy-free units. Rather, they were fully specified basic variants (cf. Anderson 1985: 43ff.). De Courtenay's view of the phoneme as an invariable unit that underlies allomorphy and allophony would become the dominant one in the twentieth century, with the further restriction that only those allomorphic variants that are phonetically (later, phonotactically) governed are taken into account (as made explicit in the Leningrad School; see §8.8.2).

8.3.2 Ferdinand de Saussure

If previous writers had already seen the important difference between allophonic differences and distinctive differences, what did Ferdinand de Saussure (1857–1913) add to this appreciation in *Cours de linguistique générale* (1916), which contains his views as written down by some of his students based on their and his lecture notes? He distinguished between the concrete sound (for which he reserved the term 'phoneme,' whose study was part of the study of *parole*, what we call today 'phonetics' but was called 'phonology' by him) and the 'acoustic image' (part of the linguistic sign's form or *signifiant* and thus falling within the realm of *langue*). For de Saussure, phonemes are contextual realizations of 'phonetic species,' and the acoustic image that he speaks of is a mental representation of this idealized unit which forms the basis for speech production and speech perception (and for alphabetic writing). Phonetic species (which are close to the modern sense of the term 'phoneme,' and which, similarly to de Courtenay, are regarded as a psychological unit) are of interest to de Saussure because they constitute the differences between the perceptible forms of linguistic signs. As such, the focus of study is not on their intrinsic or contextual properties, but rather on the relations between them, which is the essential structuralist angle that de Saussure added to the study of language and, by extension, to the study of 'sound images.' By making a distinction between the image and the actual sound, de Saussure recognized a system of implementation rules (belonging to *parole* and not to *langue*) which accounted for allophony. Like de Courtenay, his images were not composed of features and thus not completely redundancy-free, a view radically different from that of the Prague School.

Unlike de Courtenay, de Saussure's primary concern, as Anderson (1985) puts it, was representation (the phonemes and the structure of words), while de Courtenay and Kruszewski directed their primary attention to rules.

8.4 PRAGUE SCHOOL PHONOLOGY (N. S. TRUBETZKOY AND ROMAN JAKOBSON)

While the ideas of Baudouin de Courtenay (and Kruszewski) were pursued by his students in St Petersburg (and other Russian universities where he taught), a group of young linguists in Moscow worked on integrating the Kazan views and de Saussure's work into a new approach to the study of language. Several of these joined the Linguistic Circle of Prague, which, within an interdisciplinary intellectual climate that profited from both native and imported traditions and scholars, gave rise to the prolific Prague School which made its new approach to phonology as the centrepiece of the proper study of language, Nikolaj S. Trubetzkoy (1890–1938) and Roman Jakobson (1896–1982) being its most important members.[7]

The Prague School linguists felt that (a) phonology (and language in general), both from a synchronic and diachronic view point, needed to be studied in terms of systems of interrelated phonemes, rather than as units in isolation (like the nineteenth-century philologists); (b) that the focus should be on distinctive properties; and (c) that phonologists needed to formulate universal laws that govern sound systems. Jakobson wrote (and Trubetzkoy and Sergei Karcevskij, 1884–1955, co-signed) a set of theses presented at the First International Congress of Linguists, held in the Hague in 1928, which set a new direction for linguistic and specifically phonological research. Systems should be studied in terms of recurrent correlations, pointing to distinctive properties among sets of phonemes, such as voicing, and noting that these properties need to be understood as acoustic events. Here, Jakobson laid the foundation for the theory of distinctive features which he developed later.

The Prague School saw phonetics as entirely different from phonology. This strict, rather programmatic, separation was firmly embraced by some followers of the Prague School (such as Louis Hjemslev), but by no means by all. Various later proponents of this school took a more integrative approach.

The Prague School views were well advocated and dispersed, attracting notable followers elsewhere in Europe throughout the 1930s such as André Martinet, 1949) and Nicolas van Wijk, 1939). Just prior to his death, Trubetzkoy completed his *Grundzüge der Phonologie*, published in 1939, which is the most complete and fullest statement of the Prague School programme.

Trubetzkoy is very much focused on establishing the set of phonetic properties which can serve as a contrast between phonemes. In terms of these properties, phonemes form systems, which are governed by universal laws. The system was the important part of the analysis, and phonemes were primarily seen as points in this system which was a network of oppositions. If a phonological property corresponds to

---

[7] Works on Prague phonology are Vachek (1964) and Krámský (1974). A more general work on the Prague School is Toman (1995).

several phonetic properties, these, while different, cannot be distinctive. The phoneme is the sum of the phonologically relevant properties. The phoneme was first seen as a psychological unit, but for Trubetzkoy it later became a functional unit, a term in a phonological opposition. It forms part of language as a social system.

Since phonemes are specified for properties that are strictly contrastive, a phoneme that phonetically is a stop is only phonologically seen as a stop if there is another phoneme identical in all respects except for not being a stop. Although this comes close to speaking about phonemes in terms of distinctive features, Trubetzkoy did not think of the non-contrastive unit in question as underspecified, but rather as a complete phoneme. Thus, although Trubetzkoy did not speak of features as such, he proposed various classifications of phonological oppositions, one being a three-way distinction in privative (e.g. voicing, nasality), equipollent (e.g. front/back), and gradual oppositions (e.g., vowel height). Trubetzkoy also made a specific proposal for a set of suprasegmental (non-segmental) properties such as tone and accent.

A very important aspect of Trubetzkoy's phonological theory concerns the fact that there are positions in the word where oppositions are *neutralized*. A notorious case is the neutralization of the voice contrast among obstruents in final position in German. In final position we find the voiceless sound, but since there is no contrast, the underlying phonological unit has no voicing property at all. Trubetzkoy calls such a unit an 'archiphoneme.' The phonetic value that an archiphoneme acquires is the 'logically unmarked' value. In the case at hand, voicelessness is thus logically unmarked. It is, at the same time, also 'naturally unmarked' since Trubetzkoy thinks of voicing as constituting a privative opposition. In principle, the former can exist without the latter. If a gradual opposition between mid and high vowels is neutralized such that only high vowels occur in some position, high vowels would be logically, but not naturally, unmarked.

A characteristic trait of the Prague School, following from its emphasis on systems, is its typological orientation, providing many typologies of vocalic and consonantal systems.

After Trubetzkoy's death in 1938, Jakobson focused his attention on the development of a theory of distinctive features, breaking away from de Saussure's idea that the linguistic sign can only be divided sequentially and thus from the idea that phonemes are the smallest building blocks of phonological structure. He proposed to analyse *all* phonological oppositions (privative, equipollent, and gradual) in terms of binary features, which gave rise to the 'feature matrix.' For Jakobson, features were clearly building blocks and thus the true primitives of phonology; see Jakobson et al. (1952) and Jakobson and Halle (1956). In this he differed from Trubetzkoy, who had seen his phonological properties as *attributes* of phonemes.

In his choice of a rigid binary system, Jakobson reflected a deep influence from new theories of communication which had focused on the most economical transmission of information; see Cherry et al. (1952) and Goldsmith (2000).

Most characteristic of his approach was the goal to subsume related phonetic properties that were thought not to occur contrastively in any language under a single

feature, which implied the notion that there could be a single universal set of features for both consonants and vowels. He thus achieved a great reduction in his system when compared to Trubetzkoy's. Jakobson achieved this reduction by emphasizing the auditory characterization of features so that different articulatory actions that have similar auditory effects can be joined into one feature. Jakobson did adopt Trubetzkoy's view that phonemes needed to be represented minimally, such that only properties that are strictly necessary to distinguish words from each other would be specified.

Jakobson also broke new ground by trying to discover phonological universals and taking into account not just evidence from normal speech but also deviant speech, developmental speech, and historical change. His *Kindersprache, Aphasie und allgemeine Lautgesetze* (1941) brings together facts from three domains into the proposal that there is an order in the utilization of the features, meaning that certain types of contrast take precedence (in language development and in languages as such) over others, which reveals the cognitive language development, rather than the development of the articulatory motoric system. He linked these stages to implicational universals and to processes of language change.

Roman Jakobson, and his student Morris Halle, form a transition from the Prague School approaches toward those of *SPE, The Sound Pattern of English* (Chomsky and Halle 1968), discussed in §8.9.

8.5 GLOSSEMATICS (LOUIS HJEMSLEV AND HANS JØRGEN UDALL)

Establishing a new discipline often involves aggressive separation, and emphasis on autonomy and depreciation of those that are outside the new field. As Durand and Laks (2002) show, some phonologists went further than seeing phonology as being more abstract than phonetics (generalizing over non-distinctive variants) and declared that phonology is completely and logically independent from phonetic substance. Although it is not obvious how in practice one could postulate phonological units without considering the substance first, it is of course true that phonology deals with entities that are not themselves speech sounds, but instead symbolic representations of speech sounds. The idea that phonological entities are not defined or definable in terms of phonetic substance is inherent to the structuralist programme which defines phonemes as points in a system of oppositions. But the most explicit on the strict separation between phonology (the study of form) and phonetics (the study of substance) was the Glossematic School of Louis Hjemslev (1899–1965) and Hans Jørgen Udall (1907–57), who proclaimed that phonological units are purely abstract and substance-free. (The same can be said of Šaumjan's model discussed in §8.8.2, which was inspired by glossematics.) In trying to study the relation between sound (and meaning) in terms of a level of analysis which had to be conceived without relying on either of these

substances, Hjelmslev went against the empiricist nature of, for example, the American structuralists who base their analysis in the study of what can be perceived (i.e. sounds) alone. However, in practice, both approaches, of course, have to depart from the phonetic substance. The crucial difference between glossematics and the empiricist approach is that in the latter the analysis has to be motivated in terms of the phonetic substance that it is based on, whereas in glossematics this is not so. The motivation for the form analysis has to be entirely internal to the form dimension.

8.6 THE LONDON SCHOOL (DANIEL JONES AND J. R. FIRTH)

The phoneticians from the London School adopted the phoneme, because they needed this concept to cope with the fact that speech sounds occurred in a seemingly unlimited variety. Their use of this term was influenced by the Kazan School views. They did not see the discovery of this unit as the beginning of a new discipline, though. Rather, they continued to think of what they did as phonetics. Views in Great Britain built on a long tradition of precise work on the properties of speech sounds with specific reference to English, fuelled by practical rather than theoretical purposes. Thus, these views developed without much influence from older and newer continental or American approaches and in turn did not impact the development of phonology in these other parts of the world.

Alexander Melville Bell (1819–1905) developed a precise notation ('Visible Speech,' 1867) for the articulation of sounds in any language, which he hoped would be helpful in teaching the pronunciation of English and other languages (to, among others, deaf people). This approach anticipated the development of theories of subsegmental structure (see Halle 1983, 2005), although, as pointed out in §8.2.1, many earlier pheneticians had done so too. Henry Sweet (1845–1912) wrote a *Handbook of Phonetics* (1877) in which he advocated the distinction between broad and narrow transcription (in agreement with Paul Passy), the former aiming to have separate symbols only for contrastive segments. In doing so he adhered to the phonemic principle without using the term. Anderson (1985: 173) makes the important point that the mere recognition of distinctive speech sounds ('phonemes') does not put Sweet (and others before him who had done the same) in the same realm as de Saussure and the Prague School phonologists. These latter scholars not only recognized the phoneme as a contrastive unit but also, and perhaps more importantly, placed the phoneme in a coherent and organized system that is governed by recurrent use of oppositions (or distinctive features). Thus Sweet should perhaps be better seen as the founding father of phonetics as an independent academic discipline rather than as a founding father of phonology. Put differently, while the continental and American structuralists put great emphasis on theories about how to establish phonemes, others, such as the British phoneticians and indeed the

earlier students of sound change, postulated or assumed the phoneme for practical purposes (such as writing or transcription) or simply because they took it for granted that allophonic variations were to be ignored.

8.6.1 Daniel Jones

Daniel Jones (1881–1967) learned from Henry Sweet and Paul Passy that phonetic transcription needed to focus on 'broad transcriptions,' and when he later encountered Baudouin de Courtenay's work (via Lev V. Ščerba) he fully came to appreciate the phoneme unit and its practical value for spelling design and language teaching; this was before he learned of de Saussure's work or the Prague School. For Jones, the phoneme was not a psychological but a physical unit, a family of related sounds, including the free variants in each position, as well as the variants that are in complementary distribution (Jones 1929). He makes no reference to meaning and thus no reference to distinctive function. The distinctive function, he says, is what phonemes *do*, not what they *are*. His major work on the phoneme (1950)[8] appeared very late in his career. It contains detailed analyses of a wide variety of languages with a keen eye for phonetic detail. As in the case of Henry Sweet, we cannot say that Jones developed a phonological theory. His focus was on phoneme inventories. He did not focus on systematic relations between phonemes in terms of systems or in terms of alternations, nor did he pay much attention to distributional restrictions. A noteworthy property of his view of the phoneme is that he excludes the possibility of one sound in some context being the realization of two different phonemes. Thus the sound [t] alternating with [d] as a result of final devoicing could not be analysed by him as belonging to the phoneme /d/; it has to be assigned to the phoneme /t/. His views on neutralization and phonemic overlap are thus very different from those of the Prague School (but not so different from those of Bloomfield and his followers). Another difference from the Prague School is that Jones only used the term 'phoneme' for segmental units, introducing other terms (e.g. 'chronemes,' 'tonemes,' and 'stronemes') for length, tones, and stress.

8.6.2 John Rupert Firth

Unlike Jones, John Rupert Firth (1890–1960) was interested in developing a theory of language. His views on phonology did not, however, see phonemes as the foundational units (though he did acknowledge them as a sound basis for writing systems). Rather, he put great emphasis on the fact that many phonetic features belong to larger units than the single segment (which he called the 'phonematic unit'). There are systems of phonematic units for the different positions in the word and Firth (again unlike Jones

---

[8] See the 3rd edn from 1967 with an essay on the history of the term 'phoneme.'

and the continental schools) saw no need to recognize an initial [n] and a final [n] as instances of the same unit, especially not if in initial position we would have a contrast [n]–[m]–[ŋ], and no contrast (only [n]) finally. (Twaddell 1935 makes the same point; see §8.7.2.1.) This forms part of the 'polysystematic' approach, which goes much further and includes the notion that analyses of different parts of the phonology or grammar need not all be compatible with each other. Firth made no clear distinction between phonetic properties and (contrastive) phonological properties. Nor did he preclude reference to grammatical aspects in the phonological analysis. Since prosodies capture, among other things, co-occurrences of properties between segments, they capture in a static fashion what in other models (generative phonology) would be captured in rules (e.g. a rule turning an oral vowel into a nasal vowel before a nasal consonant). This, as Anderson (1985: 188) points out, precludes rule ordering. Indeed, the Firthian approach is a no-rule representational one-level theory, that can be formalized in a modern-day declarative model which also has these properties (as shown in Ogden 1999) (see §8.9.2). There is also a link with Zellig Harris' theory of 'long components' (Harris 1944) (see §8.7.2.2).

The importance of the Firthian approach, which was never fully worked out or written down (Firth 1948 being the main source), lies in its recognition of suprasegmental properties. Robins (1967), Dinneen (1967: 299–325), and Langendoen (1968) provide expositions of this approach from the viewpoint of generative grammar. An important collection of papers is Palmer (1970). Firth's approach prefigured important aspects of 'autosegmental phonology' or, more broadly, non-linear phonology which emerged in the mid-1970s in the US (see van der Hulst and Smith 1982a,b). Anderson (1985) shows how Firth's analysis of Arabic is very similar to that of McCarthy (1981), including the distinctions between a CV structure and separate vowel and consonant sequences. The prosodies, of course, resemble the autosegments proposed in Goldsmith (1976a,b) and Clements (1980 [1976], 2000). It is unfortunate that approaches and theories conceived in Great Britain tend to stay in Great Britain.

8.7 AMERICAN APPROACHES

In this section we start our overview of the American structuralism with Franz Boas, whose work of course pre-dates structuralism. Boas's interest in language was primarily in precise descriptions of Native American languages. He believed that these descriptions should record phonetic detail, and he therefore was against a broad transcription system. A precise description would have to reveal all the phonetic segments and their possible combinations. Predictabilities in phonetic properties could be recorded in a separate set of rules which capture allophony, phonemic alternations, and even the phonological/expression of semantic categories. This set of rules would imply, one might say, recognition of distinctive and predictable properties, without overt recognition of the 'phonemic principle' in the transcription itself. In this sense, Boas favoured

what one might call a phonotactic analysis of the phonetic level. Anderson (1985: 212) notes that Boas's thinking about processes contains the idea of rule ordering, albeit in a diachronic sense, when talking about historically prior phonetic forms.

8.7.1 Edward Sapir

The next grand figure is Edward Sapir (1884–1939), a student of Boas, from whom he inherited an intense interest in the relation between language and culture. Sapir sees language as a mental phenomenon in which he was, as we will see, very different from his contemporary, Leonard Bloomfield, whom we turn to in the next section. He was well aware of the Prague School ideas through his correspondence with Trubetzkoy, who in turn held Sapir's work in high esteem. With reference to phonology, his mentalistic viewpoint entailed for him that phonology was an inner system; see Sapir (1925, 1933). As such Sapir therefore almost regarded the specific phonetic properties of phonemes as epiphenomenal. This did not preclude him from being a meticulous descriptive linguist who, again like Boas, studied various Native American languages, and as such he did not disregard taking note of phonetic details. He did not think of phonemes as being composed of smaller elements, like features, but did stress that they form a system, based on their contrastive patterning, distribution, and participation in morphological alternations. In particular, Sapir's interest in sequential phoneme combinations is noteworthy, since this syntagmatic aspect of phonological structure had not been the focus of the Prague School. He also allowed that different phonemes (his basic variants) could have overlapping phonetic realizations. Phonemic representations (basic variants) were related to the phonetic varieties by a system of rules, and Anderson (1985: 236ff.) shows that Sapir's practice uses, on the one hand, rules that state regularities over the surface phonetic forms and, on the other, rules that alter basic variants to bring them in line with the surface constraints. In other words, Sapir had adopted a kind of constraint and repair system (as found in some recent developments of generative phonology: see §8.9.2). Also, Sapir alluded to the notion of rule ordering, this time not merely as a diachronic concept (cf. Kenstowicz 1975).

8.7.2 Structuralism

8.7.2.1 *Leonard Bloomfield*

Leonard Bloomfield (1887–1949) was a pivotal figure in American linguistics.[9] Through his book *Language* (1933) and his teaching he influenced a whole generation of structuralist linguists, who, despite the fact that they were all strong individual figures

---

[9] See also §18.2 below.

with their own ideas, share a common core of assumptions that can be traced back to his influence. Of specific importance was his desire to turn linguistics into an empirical science, using exact methods focusing on the observable. He also stressed that linguistics had to be independent of other fields of science (such as psychology and physics). In this empirical stance, Bloomfield was influenced by behaviourism, which was promoted by John B. Watson. This meant that meaning was defined in term of the context in which utterances were produced, rather than in terms of mental concepts. But as such, meaning did not belong to language proper. As a consequence, meaning was to play no role in the analysis of language, except in the marginal sense that to establish phonemic contrast, two speech events should differ in meaning.

Bloomfield himself did not, in fact, write a great deal on phonology, and his work on Menomini phonology, apart from a short paper in 1939, was only published in 1962, after his death (see Bever 1963). He was well aware of the trends in Europe, both neogrammarian and structuralist, and with Pāṇini's work on Sanskrit (Bloomfield 1927). He shared with the Prague School the belief that phonemes (which he did not see as being composed of features) were units that abstract away from predictable properties. Unlike the Prague School, his focus was not on paradigmatic relations between phonemes, but on syntagmatic relations (which included some attention to syllable structure, especially in the work of later scholars like Pike; see below). Phonemes were identified in terms of their combinatorial properties (an influence from Sapir). Bloomfieldians saw phonology as comprising phonemics (the study of distinctive units) and phonetics (the study of the phonetic realization). This does not mean that phonemics and phonetics are seen as independent activities: rather, phonemics is derived from phonetics. At the same time, phonetics was not really seen as part of linguistics *per se*.

Bloomfield made a clear distinction between phonetic variations of phonemes and alternations between phonemes (mostly referred to as morphophonemics by the (post-)Bloomfieldians). Alternations do no enter into the identification of phonemes, because that would violate the hypothesis that phonemes must be derivable from local phonetic properties of words, not from comparing different words. Rather, they are treated as post-phonemic, i.e. as part of morphology, and within the class of morphophonemic rules various subclasses would be distinguished, much along the lines of the Kazan School and Trubetzkoy's work on alternations. Bloomfield explicitly introduced the concept of an underlying form (or basic alternant) and the idea of ordered rules which derive the surface form, although he did not attribute a psychological reality to such analyses (see Goldsmith 2008 and Kenstowicz 1975 for discussions of rule ordering in the pre-generative period). It did not escape Bloomfield's attention that the underlying form might be similar to a historically earlier form, but he explicitly states that the underlying form and the ordered rules are part of the synchronic description. Clearly, Bloomfield's conception of morphophonemics influenced generative phonology, with the difference that in this latter model morphophonemics is seen as the core of the phonological component (see §§8.9 and 8.11).

8.7.2.2   *The post-Bloomfieldians*

A sizeable number of American structuralists developed Bloomfield's approach to phonology, which indeed left much room for elaboration, and as a consequence there are many differences in the views of the post-Bloomfieldians, among which we also find former students of Sapir. Despite their differences, the post-Bloomfieldians formed a group that developed a new standard and a common language to discuss the analysis of language; all this came with an emancipation of linguistics as an autonomous academic discipline in the US, where formerly linguistics had been regarded as a branch of anthropology.

Initially, in the 1930s, the focus of discussion was on the nature of the phoneme and the manner in which to establish phoneme inventories, with important contributions by Morris Swadesh (1909–67) and W. Freeman Twaddell (1906–82); see Swadesh (1934, 1935), contained in the Makkai (1972) collection of many important works from this period. Twaddell (1935) discusses various understandings of the phoneme, psychological (as for de Courtenay and Sapir) or physical (as for Jones and Bloomfield). He himself decides on an 'instrumental approach' (as did Zellig Harris later). With reference to the manner in which phonemes can be established, several linguists developed elaborate schemes to minimize reference to even the notion of meaning difference. Other criteria such as phonetic similarity and pattern congruity were also widely discussed. Twaddell's own views were not followed, and the Post-Bloomfieldian linguists mostly settled on a view of the phoneme as a class of non-contrastive sounds (with phonetic similarity and complementary distribution being the key criteria), much as Daniel Jones had done (see §8.6.1).

Various phonologists produced detailed analyses of the phoneme system of English, e.g. Bloch and Trager (1942), Trager and Smith (1951). Other very well-known linguists of this generation are Zellig S. Harris (1909–92), Charles F. Hockett (1916–2000), and Kenneth L. Pike (1912–2000). Harris' important contribution was his discussion of 'long components,' which resembles Firth's prosodies and foreshadows today's autosegments. Hockett (1955) followed the Prague School in a number of ways: in recognizing a sub-phonemic distinctive feature analysis (initially in the form of unary features, which he later exchanges for binary features), and an interest in the typology of phonemic systems and general laws governing their organization. Pike's contributions lie in his attention to fieldwork methods, the practical implication of phonemic analysis for the development of writing systems, his explicit recognition and development of the unit syllable, and his detailed work on tone and intonation (see Pike 1947a, b, 1943, Pike and Pike 1947). He was in several ways a critic of the prevailing structuralist views, being more open to the mixing of levels and the reference to meaning. He also developed a multidimensional view in which language utterances would have three hierarchies, the phonological, the lexical, and the grammatical, with the phoneme, morpheme, and tagmeme ('word') as its basic units. This, of course, prefigures the distinction between a prosodic and morphosyntactic partitioning of sentences that we see today in generative phonology (see §8.9.2).

The post-Bloomfieldians favoured a strict inductive approach to language analysis, aiming at pure description and staying away from trying to find explanations. From this we can understand their emphasis on *discovery procedures* which would allow the linguist to motivate his analysis starting from the phonetic signals and proceeding step by step to higher levels of analysis.

A big difference between the post-Bloomfieldian approach and generative grammar lies in the switch from this inductive bottom-up approach (phonetics > phonemics > grammar) to a deductive top-down approach (grammar > phonemics (phonology) > phonetics) which means that in generative grammar, which permits mixing levels, phonology can make reference to grammatical information.[10] Furthermore, generative grammar no longer postulates the *bi-uniqueness condition* according to which the phonemic units and (sets of) the phonetic units stand in a one-to-one relationship. Neither the thesis of separation of levels nor the prohibition of phonemic overlap was part of Bloomfield's views, but rather became typical of post-Bloomfieldians (see Bloch 1941, 1948). Another difference between Bloomfield's approach (with underlying form and ordered rules) and that of his followers signals a shift from an 'item-and-process analysis' to an 'item-and-arrangement analysis' (Hockett 1954). Bloomfield's item-and-process model was recaptured by the generative phonologists who reacted most fiercely to the post-Bloomfieldian doctrines and practices, while not fully acknowledging their own debt to Bloomfield's work.

8.8 DEVELOPMENTS IN EUROPE

Meanwhile, one might wonder what was happening in the old world after the Prague School had made its impact in the 1930s and 1940s. This section offers some brief remarks on the development in continental western Europe and in the Soviet Union, which were influenced by major political turbulences such as the Russian Revolution and the Second World War.

8.8.1 Western Europe

Most phonologists in various countries in continental western Europe continued in the footsteps of de Saussure and the Prague School. Several important contributions to the development of the structuralist approach were made by André Martinet (1908–99), specifically in the area of relating phonemic systems to patterns of phonological

---

[10] A strict adherence to the bottom-up view did lead to the postulation of so-called juncture phonemes which in fact encoded grammatical information into the phonemic representation, but this was in itself problematic, in the sense that such juncture elements would often not be directly deducible from the phonetic signal. See Scheer (2011) for extensive discussion of these issues.

change. He stressed concepts like functional load, distance between phonemes (dispersion), and systematic 'harmony resulting from economy' (i.e. maximal use of features), and the opposing force due to the asymmetry of the articulatory organs.

On the whole, there was no other specific new theoretical development or major methodological shift. Rather, we see many discussions on important themes regarding the various dichotomies that de Saussure had set up. Fischer-Jørgensen (1975: ch.12) points out that several important works were written (by people who were not always associated with a clearly defined school) that examined the relationship between *langue* (phonology) and *parole* (phonetics). Authors like Coseriu (1952), Malmberg (1964), Gunnar Fant (specifically on distinctive features), and Martinet favoured a less radical separation of these two activities. Another theme regards the parallelisms between the content (meaning) and expression planes, which were emphasized by Hjelmslev. Interesting in this respect is an analogy seen by Jerzy Kuryłowicz (1895–1978) between the syllable and the sentence (Kuryłowicz 1948). See also Malmberg (1972) on hierarchical structure in both phonology and syntax and Haugen (1956) for important work on the syllable. It would seem that a close examination of phonological activity in Europe from the 1930s to the 1960s, up to the emergence of generative phonology, is missing.

8.8.2 Soviet Union

Both Fischer-Jørgensen (1975: ch.11) and Kortlandt (1972) provide detailed overviews of phonology in the Soviet Union on which this section is based. Two schools dominate the linguistic scene, the Leningrad School and the Moscow School, both developing ideas of Baudouin de Courtenay. Fischer-Jørgensen (1975) also mentions that until the Russian Revolution there were extensive contacts between scholars in Russia and the rest of Europe. The leading scholar in the Leningrad School was Lev Vladimirovič Ščerba (1880–1944), a student of Baudouin de Courtenay, who was well acquainted with de Saussure's work. This school focused its attention not on alternations but on the nature of the phoneme, adopting de Courtenay's psychological stance (phoneme as sound image) and stressing its communicative (i.e. distinctive) function. The idea that phonemes unite sounds that are in complementary distribution was adopted, but the concept of neutralization was not, which meant that this school embraced the 'once a phoneme always a phoneme' principle (like Jones and the post-Bloomfieldians). After Ščerba's death, Lev Rafailovich Zinder took the leading part in this school. In the 1920s and 1930s the Moscow School rose to prominence, opposing the doctrines of the Leningrad School. This school took inspiration from de Courtenay's approach to alternations, although they considered only phonetically conditioned alternations as important. For them this made neutralization a cornerstone concept. Both schools did not want to sharply separate phonology and phonetics. The controversy between these two schools went on, although various attempts were made in the late 1940s to bridge the gap, among others by S. I. Bernstein.

On the whole, the 1920s, 1930s, and 1940s were dominated by suppressive tendencies due to the influence of the politically correct linguist Nicholas Jakovlevich Marr (1864–1934) and his followers, until, due to an intervention of Stalin in the early 1950s, Marr's ideas were rejected, opening up possibilities for the development of new ideas (although Western structuralism was still considered degenerate). This created an opportunity for Sebastian K. Šaumjan to break with both schools and establish a completely new approach which essentially neglected (except for criticizing it) all previous work in the Soviet Union and built instead (despite this being controversial) on Western structuralism, as well as on modern logic and cybernetics. Šaumjan proposes his two-level theory of phonology that separates the level of observation from the level of constructs, these levels being related by rules of correspondence. This idea, of course, recaptures the principled distinction between phonology and phonetics, which was stressed by Hjelmslev. Šaumjan's 1962 book in Russian was translated in 1968 as *Problems of Theoretical Phonology*. In his work on syllable structure he makes comparisons between this unit and the structure of sentences (as Kuryłowicz and Malmberg had done before him). During the 1960s more and more Western structuralist theories became available in translations which influenced scholars. Fischer-Jørgensen (1975) notes that generative grammar has not been very influential. On the whole, Soviet phonology takes a great interest in mathematical models and formalized description (see Kortlandt 1972), but this interest pre-dates the emergence of Chomsky's work, being rooted in cybernetics and information theory of the 1940s and 1950s.

8.9   GENERATIVE PHONOLOGY[11]

8.9.1   Early Developments

Generative phonology broke with the post-Bloomfieldian dogma that phonological analysis had to be based on the local information available in specific utterances. It firmly rejected the prohibition on mixing levels and the bi-uniqueness condition (see Chomsky and Halle 1965). Most notorious is the elimination of the phonemic level as a necessary step in between the morphophonemic level and the surface phonetic output. The postulation of this level entailed circumstances in which the same generalization needs to be stated twice. This occurs when, for example, a process is neutralizing in some cases but allophonic in others. Halle (1959) points out that this loss of generalization is undesirable. This argument has often been referred to as involving the abandonment of the phoneme, but this is incorrect. By eliminating the distinction between the morphophonemic level and the phonemic level, we end up with one level that is phonemic in the sense that it encodes the properties of segments that are distinctive,

---

[11] This section is partly based on van der Hulst (2004).

abstracting away from all contextually predictable properties. See Anderson (2000) for a discussion of the history of this argument and various other aspects of the manner in which generative phonology settled in. See also Encrevé (1997, 2000) for a critical evaluation of the manner in which Generative Phonology presented itself as breaking with the preceding structuralist tradition, while perhaps understating the continuities with earlier work in American structuralism, especially with Bloomfield's work, but (as shown in Goldsmith 2008) also with some of the post-Bloomfieldians.

Generative Phonology embraced the distinctive feature theory developed by Jakobson (with the modification that features were now seen as primarily articulatory units), and combined it with a model that was very similar to Bloomfield's early ideas, in which alternations were treated in terms of a single underlying form and surface forms were derived by ordered rules. In contrast to Bloomfield, though, this derivational model was not seen as a mere analytic tool but as a realistic model for how language users process language.

The phonological theory developed in the early days of generative phonology, culminating in Chomsky and Halle's (1968) *The Sound Pattern of English* (*SPE*) focused mostly on the derivational aspect, i.e. on rule format, rule application, and rule ordering; see also Halle (1962) and Chomsky (1967) for earlier statements. With respect to the representational aspect, *SPE*'s theory was deliberately minimal: a phonological representation was a linear sequence of unordered and unstructured feature bundles, provided with morphosyntactic bracketing and boundary symbols. The *SPE* system was also minimal in the sense of recognizing only two levels and one rule type that mediates between them, which was essentially a transformational rule type, like the so-called transformations in syntactic theory. The complexity of the rule system resulted from two factors. Firstly, since few restrictions were imposed on the *rule format*, the rules could get quite complicated, especially since various notational conventions allowed collapsing seemingly independent rules. Secondly, since rules could be extrinsically ordered and no restrictions were imposed on the distance between input and output, derivations could get quite long and underlying forms quite remote from the surface.

It can be said that the explanatory goal of *SPE* was to relate as many surface forms as possible, where 'relating' means 'deriving from the same input form.' Hence, with rules that could *do* anything and input forms that could *be* anything, only 'poverty of imagination' stood in the way of deriving pater*nal* (minus suffix) and *father* from the same input source (cf. Lightner 1972). The absence of a morphological theory that could place formal and *semantic* limits on the notion of relatedness stimulated the creative quest for rather abstract 'common' sources. Clearly, with so much freedom, chances to arrive at real explanatory accounts diminished in inverse proportion to the depth of the derivations that were proudly proposed; see Dresher (2005) and Durand (2006) for assessments of the achievements of Generative Phonology.

8.9.2 Post-*SPE* developments[12]

The dialectal back and forth that Anderson (1985) sees in the development of phonology at large repeats itself *within* the history of generative phonology. While *SPE* devotes considerable space to a motivation of a set of distinctive features and various issues arising with redundancy and feature specification, it also proposes an explicit model of phonological derivations, and it was this derivational aspect that led to much discussion and criticism during the seventies.

The mentalistic side of the *SPE* model was also criticized in its own right, quite independently from the kinds of formal considerations that we discussed above; see Derwing (1973) and Linell (1979) for critical assessments. Overall, experimental work offered little confirmation of the derivational aspects of *SPE* analyses that the more concrete approaches (such as Natural Generative Phonology) were seeking to dismantle. In other words, experimental research and concerns with the restrictiveness of theory were going in the same direction. The continuing trend to doubt the psychological validity of *SPE* led to further developments.

Meanwhile, proponents of the *SPE* approach, having channelled some of the discussions concerning rule ordering and depth of derivations into the development of lexical phonology (Kiparsky 1982, 1985), had shifted their attention to the representational properties of the *SPE*-model. As of the mid–late 1970s and continuing during a good deal of the 1980s, a flow of new ideas concerning various aspects of phonological representations started dominating the phonological scene (cf. van der Hulst and Smith 1982a,b,c). The incentive for some of these developments came from the rejection of *SPE*'s ban on syllable structure (criticized in Vennemann 1971 and Fudge 1969), as well as, to some extent, from pre-*SPE* models that had argued for a parallel or syntagmatic organization in phonological representations, alongside vertical, syntagmatic organization (cf. Firth's prosodic analysis, 1948; Harris's long components, 1944).[13]

The 1990s were dominated by the approach called Optimality Theory (OT), a nonderivational, constraint-based approach to phonology. At first sight, it may seem odd to introduce a non-derivational theory as the main player in a decade that is, according to the prediction of the dialectic model, supposed to focus on derivational issues. However, OT is a theory about the relationship between lexical forms and output forms, and in that sense it concerns the derivational side of the theory. The above reference to 'non-derivational' as a property of OT refers to the fact that OT does not recognize or need so-called intermediate levels, as a consequence of there being no extrinsically ordered rules or sub-components (such as a lexical and post-lexical component).

---

[12] In this chapter I will not review the developments within generative phonology in great detail. I refer to van der Hulst (1979, 2004), van der Hulst and Smith (1982c, 1985b), and, in particular, to Scheer (2011).

[13] Dinnsen (1979) offers a collection of papers which reflect the diversity of approaches that were part of the generative enterprise by the mid-1970s.

The history of constraints in phonology does not start with OT, however. Constraint-based phonologies (or proposals moving in that direction) have been around for a long time. For a historical perspective on constraints in phonology, see Paradis and LaCharité (1993) and Bird (1995). OT instantiates a particular version of this approach with the specific property that constraints are violable (or 'soft'). This softness of constraints results from the possibility of imposing an extrinsic ordering on the constraints, allowing them to be violated in the output if higher-ranked constraints enforce this.

What are the current trends? On the whole, it would seem that the idea that one can study phonology without any consideration of substance is not considered tenable (if it ever was), just as an exclusive focus on substance would make phonetic research aimless. It is generally recognized that a scientific investigation of the speech chain must recognize several levels of analysis, which, although perhaps not strictly separated, give room to more discrete symbolic phonological units and structures as well as to gradient units and structures, including various types of processes at and transitional processes between levels. As stated in Cohn (2011), this is simply what the cognitive science of speech comprises. There is nothing wrong with some researchers being focused on certain distinct properties of certain phases of this chain, while others focus on the correspondences and transitions between phases.

I conclude this section with a remark on the relevance of sign languages.[14] In Stokoe (1960) it is proposed that the form of signs in ASL can be decomposed into meaningless building blocks. Stokoe's primary motive was thus to be able to design an alphabet for writing down signs, not in the first instance as a writing system for practical use by users of sign languages, but rather as an equivalent to the IPA system for spoken languages. Henceforth phonetics and phonology thus must be understood as the study of the *form* of linguistic signs (whether their medium is sound or sight). I refer to van der Hulst (1993), Brentari (1998), and Lillo-Martin and Sandler (2006) for further discussions of this important line of research.

8.10 CONCLUSION

There are two fundamental motivations for recognizing phonemes. On the one hand, phonemes are necessary as abstractions over sets of allophones. Interestingly, it can be said that in this respect phonology (which is so fundamentally based on the notion of contrast and the *emic*) emerged in contrast to (or special focus of) phonetics (the *etic*). This, at least, seems to be the case for the development of de Saussure's views and the view of the British phoneticians. The second motivation for phonemes lies in alternations, in allomorphy, which de Courtenay and Kruszewski focused on, and here it can

---

[14] See Ch. 3 above.

be said that phonology emerged as a separate focus of morphology. Various scholars in the history of phonology associated mainly with one or the other aspect of phonology, but some schools, specifically the Prague School and Generative Phonology, united both aspects into one theory, the former keeping a clear distinction between them, the latter merging both into one format. A feature common to both motivations is to move beyond the observable properties of words to underlying, cognitive structures as well as the communicative function of language.

Throughout the history of linguistics, we see how developments in phonology lay the foundation for approaches to other aspects of grammar. This is evident in the Prague movement, American structuralism, and even in early generative grammar, which, while being dominated by work in syntax, has its roots in dealing with alternations in terms of underlying forms and transformational rules that alter these into surface forms in a series of steps. Today, at least in generative grammar, phonology has been downgraded to an evolutionary afterthought which developed so that human thought could be externalized in observable form. It is, however, not clear whether this external-ization system is 'just phonology' rather than both syntax and phonology, while the internal system (the organization of thought) is something entirely different. It would seem that in the view of most linguists both syntax and phonology are systems for the externalization and communication of thought, which then puts these two systems on an equal footing, making it likely that both are based on similar structural principles. In this sense, phonological theories can and should continue to inspire the development of syntactic theories, as they always have.

Finally, even though there is much repetition in the literature and endless discussion on procedural issues that do not always seem meaningful in retrospect, it must be said that the founding fathers of phonology (i.e. the prominent members of the Kazan and Prague Schools), as well as Americans like Sapir and Bloomfield and some of the post-Bloomfieldians (such as Twaddell, Hockett, and Pike), did much more than laying the foundation for subsequent developments of which current generative approaches are the culmination. Unfortunately, much work in contemporary phonology is done with at best indirect knowledge or hearsay of very well-developed theories of predecessors. This, in part, is due to the fact that textbooks are usually theory-specific, while works such as Fischer-Jørgensen (1975) and Anderson (1985) are not 'required literature' and thus not widely read. Even phonological handbooks (such as Goldsmith 1995, Gold-smith et al. 2011, deLacy 2007, Kula et al. 2011) often lack sufficient historical perspec-tive. A recently published compendium, van Oostendorp et al. (2011), which offers 120 survey articles in the field of phonology, offers some hope that current and prospective students of phonology will be directed to earlier work and not just to what has been published over the previous ten years.

CHAPTER 9

A HISTORY OF SOUND SYMBOLISM

MARGARET MAGNUS

9.1 OVERVIEW

9.1.1 The Problem with Sound Symbolism

The fundamental thesis underlying the field of sound symbolism has always been controversial, because it appears to be so transparently wrong. The Sound Symbolic Hypothesis is that the meaning of a word is partially affected by its sound (or articulation). If the sound of a word affects its meaning, then you should be able to tell what a word means just by hearing it. There should be only one language. In spite of this, there has always been a fairly substantial group of linguists who do not dismiss the possibility that the form of a word somehow affects its meaning. Many of those who we think of as 'great' prewar linguists (Bloomfield, Jakobson, Jespersen, Sapir, Firth) wrote works proposing that either the sound or the articulation of words has a synchronic, productive effect on their meaning.

9.1.2 Evidence

What sort of evidence have sound symbolists had for maintaining this position? Consider, for example, Lewis Carroll's *Jabberwocky*. You seem to be able to glean something from the meaning of nonsense words:

> 'Twas brillig, and the slithy toves
> Did gyre and gimble in the wabe;
> All mimsy were the borogoves,
> And the mome raths outgrabe.