Francesco Cangemi and Oliver Niebuhr

9 Rethinking reduction and canonical forms

Abstract: We conclude the book's reflections on reduction and reduced forms by exploring the complementary concept of canonical forms, which has profoundly shaped research on sound segments and their realization. Canonical forms have been described as symbolic, linear, and minimalistically contrastive representations, as in the case of phonological transcriptions of words. They have been conceived as mental word templates that can be eroded step by step in speech production, and then have to be reconstructed in speech perception. As a consequence, in theories focusing on canonical forms, reduced forms have often been relegated to energy efficiency or mere performance accidents. Drawing insights from (a) the history of linguistics (with a focus on the reasons behind the long-standing success of canonical forms) and (b) the book's contributing chapters (with a focus on how the study of reduced forms can inform linguistic theory), we identify four directions into which reduction research must be extended in the future with empirically rather than canonically defined reference forms. These are reduction patterns and reference forms in the area of prosody, reinforcement or strengthening as the antithesis of speech reduction, factors for predicting degree of reduction and their phonetic results, and, with regard to the latter, the separate contribution of reduction to communicative function. These research directions will help us to reassess our understanding of the dichotomy between canonical and reduced forms.

Keywords: canonical forms, phoneme, alphabetic writing, non-lexical meaning, non-linear representations, hypo-hyper theory

9.1 Reduced and canonical forms: A good partnership?

The contributions to this book show that so-called reduction phenomena and reduced forms are of great importance for our understanding of how speech works. This

Note: The chapter is the product of the joint effort of the two authors, but FC is mainly responsible for Sections 9.2.1–2 and 9.3.2–3, and ON for Sections 9.2.3 and 9.4.

Francesco Cangemi, IfL Phonetics, University of Cologne, Germany
Oliver Niebuhr, SDU Electrical Engineering, Mads Clausen Institute, University of Southern Denmark, Sonderborg, Denmark

https://doi.org/10.1515/9783110524178-009

applies to both research and applications in automatic speech recognition, natural text-to-speech synthesis, second language acquisition, psycholinguistics, phonetics, and phonology. Given this fact, it is surprising that reduced forms have received so little attention as yet. It is not long ago that research on reduction phenomena really got going, and it is only since recently that this line of research picked up speed.

Finding numbers in support of this statement is easy. Besides the fact that this book is the first that specifically addresses the concepts and processes of reduction, a title search from the proceedings of the International Congresses of Phonetic Sciences shows virtually no mention of these topics up to the end of the 1980s. Only a couple of titles are found for the congresses in 1991 and 1995, and then half a dozen titles for each of the following congresses – up to the latest (2015), in which the amount suddenly doubled. The numbers retrieved from the Google Scholar database are even more impressive. We found only 58 hits for "speech reduction" in the 25-year interval from 1965 to 1990. In contrast, the same interval from 1990 to present (2017) yielded almost 14 times as many hits, that is, 796. Only since January 2017, already 17 new papers dealing with "speech reduction" were added to Google Scholar. This corresponds to almost 30% of the total number of papers that Google Scholar found between 1965 and 1990. These numbers can hardly be explained by the generally increasing number of papers alone.

How could such an important aspect of speech remain "under the radar" for so long? It was stated in the introduction and reflected in virtually all preceding chapters that reduction is no simple and categorical all-or-none phenomenon but rather a complex process that manifests itself to various degrees in various conditions. Revealing this complexity, although we are still far from having the full picture, has surely contributed to make reduction a research subject in its own right that poses new challenges and calls for new joint efforts at the intersection of phonetics, phonology, and speech technology.

The hyper- and hypospeech (H&H) theory of Lindblom (1990) and the influential paper of Nolan (1992) on the "descriptive role of segments" were major driving forces for this new way of thinking. The work of Mirjam Ernestus, Natasha Warner, and colleagues, which also include the series of (Nijmegen) Speech Reduction Workshops,[1] represent valuable recent contributions to raise and maintain the awareness of the complexity of reduction. However, as strange as it may seem at first sight, it is probably not exaggerated to state that we owe the breakthrough of research on reduction to the progress in computer technology.

It was only some years ago that speech scientists still had to decide whether they wanted to make in-depth or large-scale analyses of their data (cf. Mattingly 1999).

1 http://www.u.arizona.edu/~nwarner/Workshop.html, see also Ernestus and Warner (2011).

Nowadays, computers have become so powerful that this technical barrier no longer exists. In-depth and large-scale analyses are not mutually exclusive aims anymore. As is stressed by Harrington (2010: 82): "As a result of the further development in computer technology [...] there are now large-scale acoustic databases, many of them phonetically labeled, as well as tools for their analyses," such as WebMAUS (Strunk et al. 2014), Prosomarker (Origlia and Alfano 2012), ProsodyPro (Xu 2013), and the analysis scripts of Barbosa et al. (2016). Researchers have begun to annotate and analyse the timing of face and body movements relative to linguistic elements in the acoustic domain (cf. Allwood et al. 2007), "based on speech corpora that are increasingly representative of natural, spontaneous speech" (Bird and Harrington 2001: 1). Note that it is less than 15 years ago that the outstanding speech scientist Peter Ladefoged stated in his book on Phonetic Data Analysis that cassette tapes and DAT tapes are still fairly widespread, although it is "likely that soon we will all be using some form of digital recording" (Ladefoged, 2003: 182). Ladefoged was obviously right. Yet, he underestimated how computers would change his whole discipline. An early prime example for the new opportunities offered by analyses of extensive annotated speech corpora is the study of Campbell and Mokthari (2013) on addressee-specific voice quality variation in Japanese, which was published in the same year as Ladefoged's book. About 15 years later, there is nothing "exotic" about such big-data studies anymore. In fact, six out of the seven contributed chapters in this book present data from corpus analyses, the most outstanding example being the analysis of the 900-minute Ernestus Corpus of Spontaneous Dutch (Ernestus 2000).

The reason why this development played (and still plays) a key role in impelling speech reduction research is that it provided us with one basic truth: Reduction is not an exception to the rule and hence not something that we can stay clear of in dealing with speech communication. Rather, reduction *is* the rule. Not even motherese and child-directed speech are spared from strong reductions (see Chapter 3 of van Dommelen). For example, frequency statistics derived from the Kiel Corpus of Spontaneous Speech (Wesener 1999) show that intervocalic /bdg/ are very often realized as approximants in German. In the case of /b/, like in *aber* (but), this applies to 72% of all tokens. The words *ist* (is) lacks its final /t/ in 89% of the cases; *einem* (indef. article) typically deviates more strongly from its full form and is to 80% realized as either [mm] or [m]. Liquids like /l/ are literally melting away in German spontaneous speech. About 50% of all *also* (so/also) and *mal* (discourse particle) tokens lack this sound, at least in the form of a separate segment (see the /l/ reduction in the chapter of Ernestus and Smith). Examples like these fit in well with the finding of Cutugno et al., and Adda-Decker and Lamel (Chapters 7 and 4) that so-called weak forms are particularly susceptible to reduction in German. They also fit in well with the fact that reduction is more frequent in spontaneous than in read speech (see Chapters 3, 4, and 2 of van Dommelen, Adda-Decker and Lamel, and Clopper and Turnbull).

This basic insight "that reduced forms [are] the most natural form of speech coding" (Cutugno et al., Chapter 7) obviously challenges a pervasive well-guarded concept of the speech sciences: canonical forms, that is, the symbolic, linear, and minimalistic representations of words. Therefore, we dedicate the final chapter of this book to this pervasive concept. Canonical forms and their "partnership" with phonemes and reduction offer us a privileged viewpoint to revisit, in an integrated way, some of the most important developments of linguistics in the twentieth century. On this basis, we outline with reference to important pages of the classic literature how and why canonical forms became so powerful that they were able to marginalize "troublemakers" like speech reduction for such a long time.

The roots of canonical forms can be retraced to the research questions and practices behind most work in phonology and phonetics up to the twentieth century (Section 9.2). We will see, however, that researchers from different fields and schools within the speech sciences have already anticipated some of the theoretical and empirical insights that started blooming in the recent investigations of reduced forms (Section 9.3). Following from this line of thought, we conclude that the study of reduced forms will be a catalyst in advancing research on speech communication, with reduction and canonical forms having a problematic, but not completely useless and overall improvable partnership (Section 9.4).

9.2 Why canonical forms?

In the early stages of modern linguistics, that is, from William Jones to Ferdinand de Saussure, the primacy of canonical forms can be explained on *methodological* grounds. Symbolic, linear and minimalistic representations (Section 9.2.3) were the main tool to address the questions raised in the scientific community, either on the form of written words in the early phase of diachronic studies in comparative philology or indoeuropeanistics (Section 9.2.1) or in the form of spoken words in the later phase on synchronic studies in the semiology of living languages (Section 9.2.2).

9.2.1 Indoeuropeanistics, diachrony, and written words

In what is customarily seen as the pivotal moment in the birth of comparative philology – and thus of modern linguistics (e.g. Cannon 1990; see Campbell 2007 for criticism) – Sir William Jones claimed in his speech at the Royal Asiatic Society of Bengala (1786) that "no philologer could examine [Sanskrit, Greek, and Latin], without believing them to have sprung from some common source, which, perhaps, no longer exists". This statement set the goals for generations of

indoeuropeanists in the century that followed, culminating in the groundwork by Bopp (1816) and the synthesis by Schleicher (1861–1862). Comparative philologists aimed at reconstructing a common hypothesized source language, by mapping similarities and differences between several current and extinct languages.

Crucially to our discussion, by including or focussing on extinct languages (as exemplified in Jones' quote above), indoeuropeanists had to restrict themselves to the use of *written* evidence, which is symbolic, linear, and underspecified – just as we defined canonical forms above. The use of canonical forms as the main tool of comparative philology can be illustrated with a passage from the first chapter of Saussure's *Cours*:

> a comparison of the paradigms of Latin *genus* (*genus, generis, genere, genera, generum,* etc.) and Greek (*génos, géneo, génei, génea, genéōn*, etc.) reveals nothing. But the picture changes as soon as we add the corresponding Sanskrit series (*ǵanas, ǵanasas, ǵanasi, ǵanasu, ǵanasām*, etc.). A glance reveals the similarity between the Greek forms and the Latin forms. If we accept tentatively the hypothesis that *ǵanas* represents the primitive state – and this step facilitates explanation – then we conclude that *s* must have fallen in Greek forms wherever it occurred between two vowels. Next we conclude that *s* became *r* in Latin under the same conditions. (1916: 2)

Finding a link between Sanskrit, Greek, and Latin forms thus relied, first, on isolating a unit (*s* in our example) and, second, on applying to this unit a variety of processes ("falling" or "becoming *r*"). But, while indoeuropeanists (and later on neogrammarians) put large emphasis on the study of such processes, they did not deal extensively with the concepts of units and of their isolation. After all, by working with written evidence using alphabetic scripts, constant temporally discrete units and their sequencing are almost self-evident notions (see also Section 9.3.3). This might help explain why work on the notion of phoneme, for example, only started towards the end of the nineteenth century, and with linguists (such as Kruszewski) whose interests lied more with the synchronic study of living languages than with diachronic reconstructions.

9.2.2 Semiology, synchrony, and spoken words

Canonical forms in the sense of linear sequences of discrete symbols remained the main tool also for linguists who started to focus on the synchronic aspects of living languages. In his *Cours*, Saussure redefined the goals of linguistics in the broader frame of semiology. Language is presented as "a system of signs that express ideas" (1916: 16). More precisely, linguistic signs are famously said to unite two psychological entities: "not a thing and a name, but a concept and a sound-image" (p. 66); the latter two being subsequently renamed *signifié*

(signified) and *signifiant* (signifier, p. 67). Both are psychological entities, constructs of the mind: the signified is not the object, and the signifier is not the sound either. But, what exactly are these ideas, or concepts, or *signifiés*?

In the following pages of his *Cours*, while discussing a variety of aspects about the "life" of linguistic signs, Saussure provides several examples of signs: Latin *arbor* (tree); French *sœur* (sister), and two pairs we will be focussing upon below: the signifiers of French *bœuf* and German *Ochs* (ox), and the historically related Latin *necāre* (to kill) and French *noyer* (to drown). For each of the four signs, the signified is a lexical (denotative) meaning. If the research goal is creating a system of linguistic signs, it is obvious to start with those signs for which the signified is clear-cut. Words meet this comfort criterion. In contrast, intonation and stress patterns do not; and this is not only because they convey pragmatic meanings (in Western Germanic languages) whose intricacies are by far less self-evident to the language user. In addition, it is also hard to pinpoint the syntactic and paradigmatic structure in which these stress and intonation patterns are organized (Baumann et al. 2016; Kügler et al. 2015; Ladd 2008; Torreira and Grice in press; Wagner et al. 2015). So, even if Saussure had had access to concepts such as pitch accents, edge tones, and descriptive categories such as partial topic and contrastive focus when starting the discussion on the arbitrary nature of the sign, we are confident he still would have stuck with the *bœuf*/*Ochs* example.

Crucial to our discussion, understanding words as the chief instantiation of signs makes them easier to handle not only on the side of the signified but also on the side of the signifier. Unlike the prosody of an utterance, whose representation is chronically insufficient in written language and is still largely non-consensual among researchers (Kügler et al. 2015), words seem to already come with a representation of their own. This representation is remarkably similar to its written form (with the exclusion of some "transduction" details), as exemplified in the following passage:

> The idea of "sister" is not linked by any inner relationship to the succession of sounds *s-ö-r* (*sœur*) which serves as its signifier in French; that it could be represented equally by just any other sequence is proved by differences among languages: the signified "ox" has as its signifier *b-ö-f* (*bœuf*) on one side of the border and *o-k-s* (*Ochs*) on the other. (p. 68)

These signifiers (*s-ö-r, b-ö-f, o-k-s*) inherited an important property from their written counterparts (*sœur, bœuf, Ochs*): they were conceptualized as consisting of linear sequences of discrete symbols. The influence exerted by written forms becomes evident when Saussure discusses language change. Rather than seeing language change as a consequence of either phonetic changes of the signifier or of changes in meaning of the signified, Saussure sees language change as the consequence of a shift in the relationship between signified and signifier. For example,

"Latin *necāre* (to kill) became *noyer* (drown) in French" (p. 75). But, what actually *is* the signifier, that is, the sound-image of Latin *necāre*? No doubt Saussure did have his mental representation of how *necāre* sounded. However, this imagined sound pattern had been deduced from work on written sources. That is, Saussure took for granted that grapheme sequences translate into sound sequences. This is only possible if it is beyond question that sequences of sounds are a valid concept and that these sequences always occur in the same complete (i.e. full) form.

The same passage on killing and drowning already indicates the difficulties of using written evidence for a reconstructing signs over centuries:

> If instead of comparing Classical Latin *necāre* with French *noyer*, we contrast the former term with *necare* of Vulgar Latin of the fourth or fifth century meaning "drown" the case is a little different; but here again, although there is *no appreciable change in the signifier*, there is a shift in the relationship between the idea and the sign. (p. 75, our emphasis)

The two Latin written forms only differ in the presence of a length diacritic on the vowel of the penultimate syllable. From this point of view, there is indeed no appreciable change in the signifier (at least when compared to the change in the signified). But, this point of view ignores that the loss of distinctive vowel length, and thus of the correlation between syllable weight and stress placement which operated in Classical Latin, had a massive impact on the new Latin system. Moreover, the Vulgar stress had probably nothing to do with the Classical one, both in terms of its function and its substance (cf. Burkard 2014). So, beyond what is contained in the mere chain of symbols, the signifier's "sound image" as well as its embedding in linguistic structure probably *did* change considerably over the first five centuries of the first millennium; the change from *necāre* to *necare* becomes all the more significant as we step away from the words' written representations as symbolic sequences of discrete units.

9.2.3 The phoneme

At about the beginning of the twentieth century, the consolidated use of written evidence suggested representing the signifier as a linear sequence of discrete entities, and the desire to ground semiology on intuitively accessible, tangible contrasts meant that the signified had to be a lexical meaning. In the 1930s, these two developments intertwined with the debate on the nature of the phoneme, and crystallized in the notion of distinctiveness. Phonemes were understood as discrete parts of a linear sequence, and as minimal units whose role was in the distinction between lexical meanings. By focussing on lexical distinctions, representations could be kept as minimalistic as possible, and

remained underspecified with respect to oppositions of meaning at the sentence or discourse level. This approach lies at the heart of the notion of canonical forms, which revolves around the concepts of linearity, discreteness, and (minimalistic and lexical) distinctivity.

One might wonder if what we have here is an example of the famous "cobra effect". The cobra effect means that a concept or an activity that is developed in dealing with a problem, for example, with the (implicit) intention to solve that problem, actually makes the problem worse (Siebert 2003). In a nutshell, although it is probably difficult to imagine for younger researchers, there was a time when there was no differentiation between phonetics and phonology, that is, they constituted one integrated discipline (Ohala 2004). The basic assumption in those days was that speech sounds were steady invariable units, and words and sentences were created by concatenating these units (Menzerath and de Lacerda 1933). Initial evidence of, for example, between-speaker variability and the discovery of variation due to, for example, co-articulation or the assimilation, lenition, and elision of entire speech sounds made linguistics separate these messy observations and measurements from the nicely ordered, invariable structure of language and created the need to find out how the former can be made consistent with the latter. It is obvious that the "former" is the object of phonetics, or, more specifically, experimental and instrumental phonetics, and the latter has continued as the object of the separate discipline of phonology. Not least because of the many different competences that are required to study communication, sharing the work in this way seems not a bad idea. One group of researchers takes care of the measurements and the physical nature of speech; and the other group of researchers uses these insights to develop and elaborate models and representations of sound systems.

However, what happened in practice was what Ohala (2004: 136) called the "estrangement of phonetics and phonology". That is, over time, the ties between phonetics and phonology were successively weakened or cut and that, at least in some cases, phonetic evidence was fit into phonological concepts instead of developing phonological concepts on the basis of phonetic evidence (Kohler 1995). With respect to the "cobra effect", the point is that it was, amongst other things, the discovery of reduction that favoured the rise of the phoneme. However, instead of facilitating reduction research, for instance, in the form of variant-to-category mappings, the phonemic concept, together with the overarching framework of canonical forms and lexical meanings, rather blocked or hampered studies on reduction.

Nowadays, efforts are made to bring phonetics and phonology closer together again, and studies on reduction that, for example, blurred the dividing lines between segments and prosodies as well as between co-articulation and

phonological processes (e.g. Wood 1996) can be seen as major new ties in this development. Finally, note that it was again computer technology which significantly shaped all stages of this development. It contributed to the discovery of reduction, it loosened the ties between phonology and phonetics by pushing the latter constantly further away from a traditional humanistic discipline (Ohala 2004), and, as was outlined in Section 9.1, it finally contributed to leverage reduction research and bridge the gap between phonetics and phonology.

9.3 Why not canonical forms?

While canonical forms enjoyed a lasting success in the research and teaching practices of linguists – whose usefulness as a heuristic instrument we also acknowledge – the very building blocks behind this notion were already put under scrutiny in the first half of the twentieth century. In the following sections, we briefly review three threads of work in this spirit, focussing respectively on non-lexical meaning (Section 9.3.1), on non-linear phonological representations (Section 9.3.2), and on the impact of alphabetic writing on phonemic awareness (Section 9.3.3).

9.3.1 Reduction and non-lexical meaning

We suggested above that early developments of linguistics as semiology drew the attention of researchers to meaning at the lexical level. This research agenda relegated to the background some insights on the full scope of meaning which were already well established in European culture. In this respect, it is illuminating to quote a famous passage from Dostoevskij's *A Writer's Diary* (1873–1881) on an exchange between six drunkards, revolving around a single swear word (for a recent variation on the theme, see Simon et al. 2002):

> One Sunday night I happened to walk for some fifteen paces next to a group of six drunken young workmen, and I suddenly realized that all thoughts, feelings and even a whole chain of reasoning could be expressed by that one noun, which is moreover extremely short. One young fellow said it harshly and forcefully, to express his utter contempt for whatever it was they had all been talking about. Another answered with the same noun but in a quite different tone and sense – doubting that the negative attitude of the first one was warranted. A third suddenly became incensed against the first and roughly intruded on the conversation, excitedly shouting the same noun, this time as a curse and obscenity. Here the second fellow interfered again, angry at the third, the aggressor, and restraining him, in the sense of "Now why do you have to butt in, we were discussing things quietly and here you come

> and start swearing". And he told this whole thought in one word, the same venerable word, except that he also raised his hand and put it on the third fellow's shoulder. All at once a fourth, the youngest of the group, who had kept silent till then, probably having suddenly found a solution to the original difficulty which had started the argument, raised his hand in a transport of joy and shouted ... "Eureka", do you think? "I have it"? No, not "Eureka" and not "I have it"; he repeated the same unprintable noun, one word, merely one word, but with ecstasy, in a shriek of delight – which was apparently too strong, because the sixth and the oldest, a glum-looking fellow, did not like it and cut the infantile joy of the other one short, addressing him in a sullen, exhortative bass and repeating ... yes, still the same noun, forbidden in the presence of ladies but which this time clearly meant "What are you yelling yourself hoarse for?". So, without uttering a single other word, they repeated that one beloved word six times in a row, one after another, and understood one another completely.

One might wonder whether providing an account of such an exchange is a matter that linguistics needs to deal with. Indeed, the passage above has been cited several times in the history of psychology and linguistics (e.g. Albano Leoni 2009; Vygotskij 1934; *inter alia*) to demonstrate the multifaceted nature of language and communication. Especially at Vygotskij's times, the idea of further layers of meaning besides the lexical layer was largely controversial. Jakubinskij (1923) emphasized the importance of dialogue over monologue, and Spitzer (1921) claimed the right to study all forms of linguistic exchange when publishing letters from Italian war prisoners,[2] but these were no mainstream positions. Similar concerns about the importance of meaning beyond the lexical level drove Benveniste (1974). He suggested that language should be studied not only in a "semiotic mode", but also in what he refers to as a "semantic mode", in which the view on language is widened to include the situation, the context, and the activity of the language user. This line of thought contributed eventually to the study of conversational analysis and the phonetics of talk in interaction, where neither meaning is construed in terms of a lexical network nor sound structure is understood in terms of linear discrete units (Ogden 2012, see also Section 9.3.2).

The crux of the matter is that, if canonical forms are built around lexical meaning, then, by including dialogue, discourse, and interaction in the scope of linguistic meaning, one is also questioning the validity of canonical forms as truly viable tools for the study of language. There are a number of major papers in which this issue is more or less explicitly addressed. One of them is Lindblom's

2 "The reader will perhaps find unnecessary to publish all these clumsy meaningless texts, and think one might well write down and publish coffee table conversations or fish merchants' gossips. To this I reply in Italian: *Magari*! if only the greatest possible number of everyday conversation was published! From them, psychologists and linguists could have more to learn than from their beloved written sources" (our translation).

(1990) paper on the H&H theory. In citing and outlining this seminal work, researchers often pay little or no attention to the fact that Lindblom's aim was not to explain phonetic variation in general. Rather, his theory was to explain the phonetic variation and its conditioning factors that are "required for successful lexical access" (p. 405). The key assumption of the H&H theory is that the degree of reduction is exclusively determined by the speaker's strive for articulatory economy (which includes anticipating the interlocutor's top-down processes in speech perception) on the one hand and the listener's need for sufficient discriminative power in the speech signal on the other. However, Lindblom himself stresses that this assumption of only two antagonistic forces that create the one-dimensional reduction continuum from hypo to hyper is a "deliberate simplification that is likely to be revised in the course of future work" (p. 419). Moreover, this statement about the one-dimensional simplification of H&H is made in the context of the fact that speech is "produced not only in the laboratory but also in its natural, ecological settings" (p. 418).

Lindblom's theory inspired many researchers and was refined several times, for example, by Aylett and Turk (2004; see Chapter 2 of Clopper and Turnbull). However, it has not been revised to date with respect to the simplification that Lindblom pointed out, see Niebuhr (2016) for an in-depth discussion of this fact. This is true although there is a growing body of evidence for (at least) a second dimension that drives the degree of reduction: communicative (i.e. non-lexical) meanings and functions. For example, we know for quite a long time from studies like those of de Jong (1995) and Harrington et al. (1995) that accentuation, or higher prominence levels in general, basically mean higher effort on the part of the speaker in terms of both prosodies and segments. Not so well known by now is that even more effort is put into those accents that signal new, unexpected or contrastive information (Chen et al. 2002; Dahan and Bernard 1996; Mücke and Grice 2014). In contrast, being ironic typically means investing less effort into speech production. This applies in particular to sarcastic utterances and again involves both the prosodic *and* the segmental levels (Byrant 2010; Niebuhr 2014).

Furthermore, reduction also plays a role in the syntagmatic structuring of the speech signal. For example, Local et al. (1986) and Docherty et al. (1997) showed for English that reduction variation in word-final plosives, formerly considered to be purely random, is actually systematic in that speakers reduce less at turn-final than at turn-internal phrase boundaries. Niebuhr et al. (2013) recently replicated this finding for word-final <#-en> syllables in German. Going beyond Local et al. and Docherty et al., they also conducted a perception experiment showing that listeners do in fact use the word-final segmental reduction levels to predict the end of the speaker's turn (see Graupe et al. 2014). Similarly, sounds of adjacent syllables show less strong assimilations of each other's features when there is a word boundary in

between, and again less when there is a phrase boundary in between (Kuzla 2009). Like Graupe et al. (2014), Kuzla et al. (2010) also attested the perceptual relevance of this relationship between the boundary level in the prosodic hierarchy and the degree of assimilation. In addition, regressive assimilations are much more frequent and strong than progressive assimilations, thus causing word-initial syllables to be less strongly reduced than word-final syllables, which can function as a cue to syntagmatic structure (cf. Sproat and Fujimura 1993; Vennemann 1972).

At the more social level of attitudes, speaker attributes, and pragmatic meanings, Plug (2005) analysed Dutch corpus data and found evidence for his assumption that disagreeing utterances are marked by significantly fewer and/or less strong segmental reductions. Schubotz et al. (2015) showed by means of a sub-sample of spontaneous conversations between American English speakers in the Ohio Buckeye Corpus that "discourse markers are realized with lenited segments when compared to their lexical counterparts" (p. 377) and that this reduction is stronger for younger than for older speakers. In Chapter 2, Clopper and Turnbull summarize an experiment whose result was that socio-indexical information, that is, the speaker's regional/dialectal background, is marked more strongly in the same contexts that lead to phonetic reduction. However, given Chapter 3 of van Dommelen, a similar link seems not to exist between reduction and L2 speech. Ernestus and Smith note in their chapter that reduction correlates with socio-economic status and speaker gender.

We ourselves recently conducted a perception experiment in which we varied, in three steps from canonical through moderately reduced to extremely reduced, the degree of reduction of both segments and prosodies (pitch-accent ranges and stress-induced lengthening) in a constant test sentence. The test-sentence conditions were produced by a large number of speakers so that each listener heard each reduction condition from a different speaker. Moreover, the combinations of reduction conditions and speakers were balanced across the listener sample. Listeners judged these combinations with respect to 13 different speaker attributes. Results were analysed by means of a three-way ANOVA, based on the between-subjects factors Segmental Reduction, Prosodic Reduction, and Speaker Attribute. Significant main effects of Segmental Reduction and Prosodic Reduction, as well as significant interactions of these two factors with Speaker Attribute clearly showed that segmental *and* prosodic reductions both do affect, in attribute-specific directions and orders of magnitude, how a speaker is perceived. For example, being vain was associated with less segmental but more prosodic reduction. Furthermore, unreduced canonical speech made speakers sound least tired, clumsy, and scatty, but most educated and optimistic. Sounding maximally athletic, sincere, sociable, and composed required moderate degrees rather than no or high degrees of segmental and prosodic reduction (see Niebuhr 2017).

In summary, research on reduced (or, more generally, non-canonical) forms is often most fruitful and convincing when dealing not with lexical contrasts, that is, the traditional objects of canonical forms, but with interactional and pragmatic meanings. Particularly when it comes to attitudinal meanings and speaker attributes the reduced form can actually tell *more* than the full one. We illustrate and conclude this line of argument with Hawkins' example on the use of [ə̪ə̪ə̪], a massively reduced form for English *I don't know*, which

> could allow successful communication between relaxed family members. For example, it could be said by B when A asks B where the newspaper is, and B does not know, but does not feel that she needs to stop reading her book in order to help find it. Person A should understand from this that he should not expect help in looking for the newspaper, and should either stop talking to B, or introduce a more interesting topic. (Hawkins 2003)

9.3.2 Reduction and non-linear representations

Another path that leads to questioning canonical forms comes from research on non-linear representations of sound structure. Interestingly, the idea that words contain more than a sequence of phonemes can be glimpsed even in Trubeckoj's work:

> The signifier aspect of every word in the system of a language can be analyzed into phonemes, that is, it can be represented by a particular sequence of phonemes. Of course, the matter should not be oversimplified. The phonemes should not be considered as building blocks out of which individual words are assembled. Rather, each word is a phonic entity, a *Gestalt*, and is also recognized as such by the hearer, just as an acquaintance is recognized on the street by his entire appearance [...] As a *Gestalt*, each word always contains something more than the sum of its constituents (or phonemes), namely, the principle of unity that holds the phoneme sequence together and lends individuality to a word. Yet in contrast with the individual phonemes it is not possible to localize this principle of unity within the word entity. Consequently one can say that each word can be *completely analyzed* into phonemes, that it *consists of* phonemes. (1939: 35)

In this passage, while advocating the possibility of analysing words as linear sequences of phonemes, Trubeckoj suggests that words also have a holistic silhouette (along the lines of the *cement* of Kruszewski 1883[3] and the *Klanggesicht* of Bühler 1934; see Albano Leoni 2009). This "phonetic silhouette" is also claimed

3 "A sound complex cannot be considered a mechanical juxtaposition of a certain quantity of independent sounds. When combining with one another, sounds [...] accommodate themselves

to be important in perception, but since its features cannot be consistently associated with specific sound segments within a word, it is kept out of the scope of phonology. This very consequence is rejected by research on non-linear representations of sound structure, which seeks to provide a description of such non-local features and, in doing so, necessarily questions the adequacy of an approach based on exclusively linear representations.

It is beyond the scope of this chapter to provide a detailed account of non-linear approaches to phonology (e.g. see Chapter 8 of Espy-Wilson et al. for an outline of Articulatory Phonology and the Task-Dynamic Model of Speech Production). We will merely provide pointers to the relevance of this line of research for the notions of canonical and reduced forms. In particular, the notion of *prosodies* as developed by Firth (1948) is crucial insofar as it represents a bridge between Trubeckoj's neglected concerns on the one hand and the later literature on reduction on the other. Prosodies are defined as abstractions that concur to describe "word structure and its musical attributes", thus going beyond the linear and discrete representation provided by "the total phonological complex" (Firth 1948: 123). In this sense, Firth's prosodies relate to the phonetic silhouettes that hold words together, rather than to the small coloured bricks that account for differences in the signifier:

> Let us regard the syllable as a pulse or beat, and a word or piece as a sort of bar length or grouping of pulses which bear to each other definite interrelations of length, stress, tone, quality – including voice quality and nasality. The principle to be emphasized is the *interrelation of the syllables*, what I have previously referred to as the *syntagmatic relations*, as opposed to the *paradigmatic* or *differential relations* of sounds in vowel and consonant systems, and to the paradigmatic aspect of the theory of phonemes, and to the analytic method of regarding contextual characteristics of sounds as allophones of phonematic units. (1948: 128)

Work on such non-local syntagmatic interrelations has continued challenging a strict linear-based approach, as in the case of evidence from so-called short and long domain /r/-resonances in English (Heinrich et al. 2010; Kelly and Local 1986), which are responsible for pairs such as *miller* and *mirror* having pervasive acoustic differences beyond the intervocalic material.

Prosodies in this sense are seen by Kohler (1999) as what is left of a word (or group of words) when uttered in contexts favouring hypoarticulated speech. Such "articulatory residues may persist as non-linear, suprasegmental features of syllables, reflecting, e.g., nasality or labiality that is no longer tied to specific segmental units" (Kohler 1999: 89). That is, to use a paradoxical image, under the heat of

---

to one another. This accommodation is the *cement* which transforms several sounds into one integral complex." (Kruszewski 1883: 63)

spontaneous hypoarticulated communication, it is the phonetic silhouette that remains, not the building bricks that flesh out this silhouette. The fullest account of this approach is given in Niebuhr and Kohler (2011), who define articulatory prosodies as "distinctive suprasegmental vocal-tract and phonation features that identify words in spite of segmental reduction" (p. 320). Here, prosodies are seen as constituting the "phonetic essence" of words, and thus are given a sort of ontological primacy over segments. In their follow-up paper, Kohler and Niebuhr (2011) consider the example of German *Ihnen* (dative of courtesy pronoun), which is analysed in its canonical form as [iːnən], but which is often uttered by speakers (and recognized by listeners) as [iːnn̩] or [nʲnʲ]. These reduced forms

> can be related to the same class (i.e. *Ihnen*) without an elaborate derivation from one canonical representation, because they both contain palatality and long alveolar nasality, as do other intermediate degrees of reduction. This means that all phonetic forms of this word must contain these features; they constitute the *phonetic essence* of *Ihnen*. This concept of phonetic essence may be assumed to apply to function words generally and possibly even to all lexical items. The phonetic essence of a lexical item manifests itself either in segmental units in the less reduced forms or as articulatory prosodies in more extreme reduction, where it appears to be sufficient for the listener to identify the word. (Kohler and Niebuhr 2011)

In this sense, research on reduction helps uncover the *fil rouge* that runs through Kruszewski's *cement*, Bühler's *Klanggesicht*, Trubeckoj's *principle of unity*, Firth's *prosodies*, and Niebuhr and Kohler's *phonetic essence*, which might also be glimpsed in Johnson's (2004) *islands of reliability* and Ernestus and Smith's *core properties* (Chapter 5). This connection is sometimes subterranean, sometimes explicitly acknowledged. And it ultimately joins ends with another thread of research which questions the viability of canonical forms: the issue of alphabetic writing and phonemic awareness.

9.3.3 Reduction and alphabetic writing

The implicit role of written alphabetic representation on research in phonology has surfaced several times in linguistic research throughout the course of the twentieth century. The positions of individual researchers differ in terms of the strength of the conclusions drawn, but all converge towards the need of questioning the viability of a representation of sound structure based on sequences of discrete units.

An early formulation of the problem, which joins ends with our discussion in Section 9.2.1 on the importance of written evidence in the early phases of modern linguistics, can be found again in Firth (1948):

> The development of comparative philology, and especially of phonology, also meant increased attention to transliteration and transcription in roman letters. Sir William Jones was not in any position to understand how all this might contribute to the tendency, both in historical and descriptive linguistics, to phonetic hypostatization of roman letters, and theories built on such hypostatization. In introducing my subject I began with sounds and the Roman alphabet which has determined a good deal of our phonetic thinking in western Europe. (pp. 125–126)

Alphabetic scripts, which represent words using sequences of symbols relating to units smaller than a syllable, are thus considered by Firth as the cause (and not the consequence!) of our tendency to think of words as composed of segments. The alternative view sees alphabets as a proof for the pre-existence of (psychologically or ontologically) meaningful segments. Fowler (2010: 58) discusses alphabetic writing systems, observing that "Their inventors must have had the impression that the spoken language had units to which the letters would correspond. Yet they had no alphabetic writing system to give them that impression". This notion of alphabet as an *invention*, while apparently intuitive at first sight, had actually been challenged by historical accounts of the development of writing systems, as in Gelb (1952), Sampson (1985), and, perhaps with slightly larger resonances in the linguistics research community, Faber (1992). This line of research suggests that alphabets are a *discovery*. That is, phonemic awareness is the consequence of the emergence of alphabetic systems out of the adaptation of writing systems across languages. In Faber's words, when discussing the derivation of the Greek alphabet from its Canaanite sources (in which vowels were not represented),

> it is not necessary to base an explanation for the structure of the Greek alphabet on the unattested existence of an unknown genius. The names of the Greek letters *alpha*, *beta*, etc., meaningless in Greek, have clear sources in a Canaanite acrophonic tradition, whereby each sound is associated with an object whose name begins with that sound. This fixed order of a traditional, invariant list is comparable to modern radio alphabets like *able*, *baker*, *Charlie*, etc. [...] Transmission of the Canaanite script using the acrophonic principle would have led to the *mis*interpretation of several Canaanite consonant symbols as representing vowels instead. The Canaanite words *ʔalpa* 'cow', *he* '?', *yoda* 'hand', and *ʕayna* 'eye', standing for /ʔ/, /h/, /y/, and /ʕ/, would have been perceived by speakers of a language in which, as in Greek, these sounds did not occur, as beginning in [a], [e], [i], and [ɑ], respectively. Thus, Phoenician [ʔalpa], with an initial [ʔ] became Greek [alpa], with no [ʔ]. (1992: 126)

Kohler (1995) takes an intermediate position in the debate about whether the invention of alphabetic systems is a consequence of phonemic awareness or vice versa. He claims that an alphabetic writing system has been developed only once in human language evolution, with all further systems being derived from this first system. In his opinion, the idea of representing complex sound patterns by means of sequences of discrete symbols was a direct consequence of the

three-consonant roots and their association with semantic fields in the lexicon of the Semitic language family. This morpho-phonological peculiarity favoured the emergence of mental models in which consonants and vowels were represented as separate elements, so that these elements, in turn, were finally represented by separate symbols in written language. Thus, Kohler sees phonemic awareness as a prerequisite for the development of alphabetic writing systems but stresses at the same time that phonemic awareness is not automatically created by inherent properties of linguistic structure. Rather, it needs very special conditions to occur.

We have already mentioned Firth's claim on how the Roman alphabet "has determined a good deal of our phonetic thinking in Western Europe". This fits in with the view expressed in O'Connor (1983: 441), according to which native speakers' analyses of a language are reflected (albeit not always systematically) in the structure of its orthography. Along these lines we might place evidence from Morais et al. (1979) on the fact that illiterates might have little awareness of segments, or none at all. Rather, there is increasing evidence that the syllable rather than the phone(me) is the basic unit of speech, see Greenberg (1996) for a summary with a focus on speech perception. Also famous phenomena like "phonemic restoration" are not inconsistent with this assumed primacy of the syllable. More detailed investigations confirm that restoration occurs but suggest that what listeners restore are syllabic rather than phonemic units (cf. Niebuhr 2011).

Ladefoged (1984) looks at the debate about the effects of alphabetic writing systems on spoken language representation from a meta-linguistic point of view. He states that the apparence of the Greek alphabetic system, "produced out of the spare symbols of a syllabary", set in motion a "startling conspiracy", tricking linguists into thinking that only because speech can be described in terms of segments, then language must also be structured in that way. Ladefoged elaborates on the classical example of Lindblom et al. (1984) on termite nests (see also Bybee 2001), which appear to the outside observer as having a structure revolving around pillars and arches, but are ultimately built by a simple behavioural pattern – the accumulation on grains of earth on spots on the ground containing pheromone secretions:

> Phonemes may be like arches in termite nests, visible to outside observers, but having no meaningful role in the activity of the individuals producing them. Speech *appears* to be composed of sequences of segments because of the interactions of the different systems of which it is composed. The complex gestures involved in producing syllables have diverse parts that look as if they are categorically distinct. We call these diverse parts vowels and consonants, but we must always remember that these are just names for readily distinguishable aspects of the stream of speech. Those of us who have been exposed to an alphabetic tradition may be influenced so that we are very conscious of the possibility of describing speech in terms of units of this kind. (Ladefoged 1984: 93–94)

Here we find echoes of Firth's views both on phonemic awareness as a product of alphabetical training[4] and on the polysystemic nature of sound structure (see also Hawkins 2003).

Port (2010) further elaborates on these ideas. He sees language as a "social institution that is shaped by generations of users" (much alike termite nests), which is only poorly described by segmental-based representations. These are "partially a side effect of our years of literacy education and extensive practice of literacy" (cf. Firth's "phonetic hypostatization of roman letters"). He thus rejects canonical representation of words as minimalistic strings of discrete units, and advocates richer representations. Such representations would be stored in the memory of language users not as "low-bitrate" vectors of phonemes, but as rich traces, containing detailed phonetic information which is used in our processing of linguistic variability (e.g. idiolectal and sociolectal). In doing so, Port brings together the reflexion on the "alphabetic fallacy" with research on episodic memory.

Crucial to our purposes, research on reduction has a lot to contribute to the debate on episodic memory. Johnson (2004) has convincingly shown that the traditional view of a mental lexicon built around individual word representations consisting of sequences of discrete units is inadequate to account for auditory recognition of connected speech. "Massive" reduction, as the one relating forms such as [dəvɪ̃i̯ fʌdʒ] to *divinity fudge* in Stampe (1973) example, are shown to occur frequently in spontaneous speech. Speakers' knowledge of words is not adequately represented by entries such as *divinity* = / dəvɪnəti/ in two respects – in assuming that a word is represented with a single entry, and in assuming that this entry is composed by a sequence of discrete phonemes.

9.4 Reduction and canonical forms: Assessing the partnership

Research on reduction has continuously made us question theoretical assumptions and analytical practices which were, until then, commonplace in speech sciences. Among these established assumptions or practices, on which the present chapter focused, are canonical forms. So can we conclude from what we have briefly summarized in Sections 9.2 and 9.3 that reduction and canonical

4 "We ABC people, as some Chinese have described us, are used to the process of splitting up words into letters, consonants and vowels." (Firth 1948: 122)

forms do not have a good partnership? As is typical of scientific conclusions, the answer is not that simple.

Traditional canonical forms and their basic building blocks, the phonemes, are obviously useful for those who have to abstract away from reduction phenomena in order to, for example, develop and deal with alphabetic writing systems. Also, the chapter of van Dommelen suggests that canonical forms are a practical instrument for those who need a clear and easy entry to teaching and learning languages. Moreover, van Dommelen's chapter demonstrated in accord with the chapters of Cutugno et al., Adda-Decker and Lamel, and Clopper and Turnbull that for us, the researchers, canonical forms are handy points of reference that help us detect and describe the reduced and variable sound patterns that speakers produce. In other words, traditional canonical forms have a practical advantage for all of those who *talk about* speech communication. However, in view of the accumulating evidence that is provided by this book and summarized in Section 9.3, we dare to join the voices of those who claim that traditional canonical forms are likely not as relevant for those who actually *do* speech communication (e.g. see Kohler 2000).

This need not mean that, in understanding speech communication, we should throw over board the *basic idea* of canonical forms altogether. Evidence from perceptual restoration (which may not be simply phonemic restoration, see Niebuhr 2011), phoneme monitoring, the McGurk effect (Cox et al. 1999), segmental intonation (Niebuhr 2012), and many other findings clearly show that, compared to the acoustic signal, the hearer's speech perception can well be richer in phenomenological or structural respects. In fact, alternative concepts of proper, perceptually relevant reference forms that underlie or interconnect reduction phenomena are already waiting to be further elaborated. These alternatives are not based on linear, symbolic, and minimalistic representations. Rather, they put focus on *rich phonetic detail* in the form of "gross and subtle acoustic characteristics" (Ernestus and Smith, Chapter 5) that can be either sub-phonemic or supra-phonemic in that they "may be distributed across [...] sounds" (Espy-Wilson et al., Chapter 8). The keywords that reflect these alternative reference concepts in the chapters of this book include, for example, landmarks (Cole and Shattuck-Hufnagel, Chapter 6), articulatory prosodies, core properties, and phonetic essence (Ernestus and Smith, Chapter 5).

In terms of this next generation of alternative reference concepts, the chapters of this book also gave us an idea of which major questions will drive our investigation of speech communication with respect to reduction phenomena in the future.

First, one of the major questions will be how we define reduction in the future and whether it makes sense to continue using the term at all. For example, in terms of a decrease in duration of a particular linguistic unit, as, for example, in the

chapter of Adda-Decker and Lamel, or, similarly, in terms of "less acoustic-phonetic substance", as in the chapter of Clopper and Turnbull (cf. also Espy-Wilson et al.), the term reduction is quantifiable and hence still useful and intuitive. However, reduction seems to be more and more often equated simply with "abundant phonetic variability" (Chapter 3, van Dommelen) and phonetic variation in general. Speaking of reduction in these contexts is dispensable, as the equation makes no reference to a clear-cut, superordinated reference form anymore. The definition of reduction as "fewer phonetic cues to contrastive phonological units" by Cole and Shattuck-Hufnagel is at least not generally applicable, as empirical studies showed many times now that cues bundled in the form of sound segments can be "recoded" into articulatory prosodies when the sound segments themselves have disappeared, see the chapter of Ernestus and Warner. Cutugno et al. state that "every sound segment can undergo coarticulation" and then add to this statement that coarticulation is the most widespread form of reduction. Definitions like these are similar to equating reduction to variation, but they are terminologically still more problematic as they relate reduction to individual sound segments and in this way decouple the term from its original foundation, that is, the canonical form at the word level. That is, by stating that individual sound segments are reduced, phonemes become the new canonical forms; and, unless coarticulation affects distinctive features, it is simply impossible to say whether or not a sound segment is reduced (exceptions are cardinal vowels and reduction in the sense of a decrease in duration). For example, which of the allophonic variants of German /x/ ([ç], [x], or [χ]) is reduced against which reference form, and is the English light /l/ the reduced variant of the dark /l/ or vice versa?

We think that using the term reduction is still useful. However, in line with our distinction between talking about versus actually doing speech communication, reduction is useful in the sense of a phonetic parameter rather than as basic aspect of speech cognition and representation. The concept of canonical forms is also not completely obsolete. Yet, it should be further developed to more flexible forms that take into account empirical frequencies of word variants and include, wherever necessary, multi-word expressions rather than individual words (cf. Chapter 4 of Adda-Decker and Lamel). Moreover, it should allow for a counterpart of reduction, that is, *strengthening* (cf. Chapter 6 of Cole and Shattuck-Hufnagel). In fact, it seems that the idea of strengthening as a counterpart of reduction is already in use, but not as a mature, consistently applicable concept that is firmly grounded in a revised framework of canonical forms. Rather, inspired by the H&H terminology, people currently have to use terms such as "hyperspeech" and "hyperarticulated speech" whenever they have the impression that a certain speech pattern or way of speaking exceeds the articulatory or acoustic parameter ranges that can be expected under "normal" circumstances.

Second, closely linked with the first question about the use and usefulness of reduction and canonical forms, future studies will have to deal with the question of reduction in the domain of prosody. It seems to be widely accepted that lexical stress or prominence levels can undergo reduction, see Wagner et al. (2015) and the chapters of Ernestus and Smith, and Clopper and Turnbull. This is probably because stress is quantifiable in terms of duration (a common parameter in measuring reduction anyway) and moreover closely linked to words and lexical meaning. However, what about intonation? Clopper and Turnbull, for example, found lower intonation peaks for those pitch accents that were realized in predictable focus conditions. Are lower intonation peaks instances of reduction? (In the light of the effort code of Gussenhoven 2002, this would probably be the case.) Niebuhr and Hoekstra (2015) showed in a production study that Northern Frisian speakers produce intonation plateaux rather than higher intonation peaks under expressive conditions (including contrastive focus). Is this also a case of intonational reduction? Would it be possible to define an equivalent of canonical forms for intonational units like pitch accents and edge tones? And what about reductions and/or canonical forms of voice quality and loudness? All these important theoretical and empirical questions are currently widely unaddressed.

Third, Cutugno et al. state with regard to reduction that "Spontaneous speech is characterized by a great amount of unpredictable phenomena". Predictability plays an important role in reduction research, for example, in the related fields of psycholinguistic models of word recognition and automatic speech recognition. However, note that humans probably have to rely much less on bottom-up predictability than machines when it comes to word recognition in natural everyday conversation, see Chapters 4 and 8 of Adda-Decker and Lamel, and Espy-Wilson et al.

Cutugno et al. are certainly right with their above statement about the great amount of unpredictable phenomena, especially in view of the great range of within- and between-speaker variations in the pronunciation of words under constant conditions (see Chapters 5, 6, and 8 of Ernestus and Smith, Cole and Shattuck-Hufnagel, and Espy-Wilson et al.). Nonetheless, the amount of unpredictable reduction variation should also not be overestimated, as we have only just begun to uncover the factors that determine – and hence predict – speech reduction. As was summarized in Section 9.3, differences in the degree of reduction have communicative functions, and some of the supposedly random variation became explainable and predictable on this basis. This book is a further big step in advancing our insights into the determining factors of variation in reduction. For example, Ernestus and Smith showed that the rhythmic embedding of a word influences its degree of reduction. Moreover, Clopper and Turnbull add quite a bit of complexity to the already established reduction triggers by showing that these factors are not simply additive but interact in how they affect

individual reduction parameters, and that not only reduction itself but also its triggering factors are scalable. This line of research on identifying and understanding reduction triggers and their interaction should receive special attention in the future. Computer-based analyses of spontaneous dialogues and natural everyday speech recordings will contribute a great deal to address open questions. If lab speech has to be used, we also need a better understanding of how environment and task conditions affect speech production, and whether they can even be adjusted such that they facilitate rather than impede the production of reduced forms (see Chapter 3; Niebuhr 2015).

Last but not least, another major challenge of future research will be to revise the H&H theory of Lindblom (1990) and the alternative frameworks outlined by Clopper and Turnbull such that variation in the degree of reduction is no longer a simple one-dimensional trade-off between economy and comprehensibility. This book showed clearly that reduction is much more complex and involves a lot of additional, partly antagonistic forces; and any future framework that wants to explain phonetic variation has to find a way to address this complexity; see also Niebuhr (2016). Such a model would be a great achievement, also because its value would not be confined to the humanistic fields of the speech sciences. Rather, it would also have practical implications for improving speech technology and developing materials and strategies for (second) language teaching.

In general, as research on reduction is becoming increasingly intense and dynamic, it does not take much to foresee that the number of questions and challenges posed by reduction research will further grow in the future.

References

Albano Leoni, F. 2009. *Dei suoni e dei sensi. Il volto fonico delle parole*. Bologna: il Mulino.

Allwood, J., L. Cerrato, L. Dybkjaer, K. Jokinen, C. Navarretta & P. Paggio 2007. The MUMIN multimodal coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation* 41. 273–287.

Aylett, M. & A. E. Turk 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47. 31–56.

Barbosa, P. A., Z. A. Camargo & S. Madureira 2016. Scripts for the acoustic analysis of speech data. In S. Madureira (ed.), *Sonorities: Expressivity in speech, singing, and reciting*, 164–174. São Paulo: Pontíficia Universidade Católica de São Paulo.

Baumann, S., O. Niebuhr & B. Schroeter 2016. Acoustic cues to perceived prominence levels – evidence from German spontaneous speech. Proceedings of 8th International Conference of Speech *Prosody*, 1–5. Boston, USA.

Benveniste, E. 1974. *Problèmes de linguistique générale II*. Paris: Gallimard.

Bird, S. & J. Harrington 2001. Speech annotation and corpus tools. *Speech Communication* 33. 1–4.
Bopp, F. 1816. *Über das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprache*. Frankfurt: Andreäischen Buchhandlung.
Bühler, K. 1934. *Sprachtheorie. Die Darstellungsfunktion der Sprache*. Jena: Fischer.
Burkard, T. 2014. Mythen und freie Erfindungen in der lateinischen Grammatik – das Nicht-Verstehen einer toten Sprache und seine Konsequenzen. In O. Niebuhr (ed.), *Formen des Nicht-Verstehens*, 45–76. Frankfurt: Peter Lang.
Bybee 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
Byrant, G. A. 2010. Prosodic contrasts in ironic speech. *Discourse Processes* 47. 545–566.
Campbell, N. & P. Mokhtari 2013. Voice quality: The 4th prosodic dimension. *Proceedings of 15th International Congress of Phonetic Sciences*, 2417–2420. Barcelona, Spain.
Campbell, N. 2007. Why Sir William Jones got it all wrong, or Jones' role in how to establish language families. In J. Lakarra & J. Hualde (eds.), *Studies in Basque and historical linguistics in memory of R.L. Trask*, 245–264. Bilbao: Universidad del País Vasco/Euskal Herriko Unibertsitatea.
Cannon, G. 1990. *The life and mind of oriental Jones: Sir William Jones, the father of modern linguistics*. Cambridge: Cambridge University Press.
Chen, A., C. Gussenhoven & T. Rietveld 2002. Language-specific uses of the effort code. *Proceedings of 1st International Conference of Speech Prosody 2002*, 211–214. *Aix-en-Provence, France*.
Cox, E. A., L. W. Norrix & K. P. Green 1999. The contribution of visual information to on-line sentence processing: Evidence from phoneme monitoring. *Proceedings of Audio-Visual Speech Processing Conference*, 30–36. Santa Cruz, USA.
Dahan, D. & J.-M. Bernard 1996. Interspeaker variability in emphatic accent production in French. *Language and Speech* 39. 341–374.
de Jong, K. J. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 97. 491–504.
Docherty, G. J., J. Milroy, L. Milroy & D. Walshaw 1997. Descriptive adequacy in phonology: A variationist perspective. *Journal of Linguistics* 33. 275–310.
Dostoevskij, F. 1994 [1873–1881]. *A writer's diary*. Evanston: Northwestern University Press.
Ernestus, M. 2000. *Voice assimilation and segmental reduction in Dutch*. PhD dissertation, University of Utrecht, The Netherlands.
Ernestus, M. & N. Warner 2011. An introduction to reduced pronunciation variants. *Journal of Phonetics* 39. 253–260.
Faber, A. 1992. Phonemic segmentation as epiphenomenon. In P. Downing, S. Lima & M. Nooman (eds.), *The linguistics of literacy*, 111–134. Amsterdam-Philadelphia: Benjamins.
Firth, J. R. 1948. Sounds and prosodies. *Transactions of the Philological Society* 47 (1). 127–152.
Fowler, C. 2010. The reality of phonological forms: A reply to Port. *Language Sciences* 32. 56–59.
Gelb, I. 1952. *A study of writing*. Chicago: University of Chicago Press.
Graupe, E., K. Görs & O. Niebuhr 2014. Reduktion gesprochener Sprache – Bereicherung oder Behinderung der Kommunikation? In O. Niebuhr (ed.), *Formen des Nicht-Verstehens*, 155–184. Frankfurt: Peter Lang.
Greenberg, S. 1996 Understanding speech understanding – towards a unified theory of speech perception. *Proceedings of ESCA Tutorial and Advanced Research Workshop on the Auditory Basis of Speech Perception*, 1–8. Keele, England.

Gussenhoven, C. 2002. Intonation and interpretation: phonetics and phonology. Proceedings of the First International Conference of Speech Prosody, 47–57, Aix-en-Provence, France.
Harrington, J., J. Fletcher & C. Roberts 1995. Coarticulation and the accented/unaccented distinction: Evidence from jaw movement data. *Journal of Phonetics* 23. 305–322.
Harrington, J. 2010. Acoustic Phonetics. In W. J. Hardcastle, J. Laver & F. E. Gibbon (eds.), *The handbook of phonetic sciences*, 81–129. Oxford: Wiley-Blackwell.
Hawkins, S. 2003. Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics* 31 (3–4).373–405.
Heinrich, A., Y. Flory & S. Hawkins 2010. Influence of English resonances on intelligibility of speech in noise for native English and German listeners. *Speech Communication* 52. 1038–1055.
Jakubinskij, L. P. 1979 [1923]. On verbal dialogue. *Dispositio* 4 (11–12). 321–336.
Johnson, K. 2004. Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (eds.), *Spontaneous speech: Data and analysis*, 29–54. Tokyo: The International Institute for Japanese Language.
Jones, W. 1798 [1786]. The third anniversary discourse, delivered 2d February, 1786: on the Hindus. *Asiatick Researches* 1. 415–31.
Kelly J. & J. Local 1986. Long domain resonance patterns in English. *International conference on speech input/output: Techniques and applications*, 304–309. London: Institute of Electrical Engineers.
Kohler, K. J. 1995. Phonetics – a language science in its own right? *Proceedings of 13th International Congress of Phonetic Sciences*, 10–17. Stockholm, Sweden.
Kohler, K. J. 1999. Articulatory prosodies in German reduced speech. *Proceedings of 4th International Congress of Phonetic Sciences,* 89–92. San Francisco, USA.
Kohler, K. J. 2000. Investigating unscripted speech: Implications for phonetics and phonology. *Phonetica* 57. 85–94.
Kohler, K. & O. Niebuhr 2011. On the role of articulatory prosodies in German message decoding. *Phonetica* 68. 57–87.
Kruszewski, M. 1995 [1883]. *Outline of linguistic science*. In K. Koerner (ed.), *Writings in general linguistics*, 35–173. Amsterdam-Philadelphia: Benjamins.
Kügler, F, B. Smolibocki, D. Arnold, B. Braun, S. Baumann, M. Grice, S. Jannedy, J. Michalsky, O. Niebuhr, J. Peters, S. B. Ritter, C. T. Röhr, A. Schweitzer, K. Schweitzer & P. Wagner 2015. DIMA – Annotation guidelines for German intonation. *Proceedings of 18th International Congress of Phonetic Sciences,* 1–5. Glasgow, UK.
Kuzla, C. 2009. *Prosodic structure in speech production and perception*. Nijmegen: MPI Series in Psycholinguistics.
Kuzla, C., M. Ernestus & H. Mitterer 2010. Compensation for assimilatory devoicing and prosodic structure in German fricative perception. In C. Fougeron, B. Kühnert, M. D'Imperio & N. Vallée (eds.), *Laboratory phonology* 10. 731–757. Berlin: De Gruyter.
Ladd, D. R. 2008. *Intonational Phonology*. Cambridge: CUP.
Ladefoged, P. 1984. 'Out of chaos comes order': Physiological, biological, and structural patterns in phonetics. In M. P. R. Van den Broeke & A. Cohen (eds.), *Proceedings of the 10th International Congress of Phonetic Sciences*, 83–95. Dordrecht: Foris Publications.
Ladefoged, P. 2003. *Phonetic data analysis. An introduction to fieldwork and instrumental techniques*. Malden: Blackwell.
Lindblom, B., P. MacNeilage & M. Studdert-Kennedy 1984. Self-organizing processes and the explanation of language universals. In B. Butterworth, B. Comrie & O. Dahl (eds.), *Explanations for Language Universals*, 181–203. The Hague: Mouton.

Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H and H theory. In W. Hardcastle & A. Marchal (eds.), *Speech production and speech modelling*, 403–439. Dordrecht: Kluwer.
Local, J., J. Kelly & W. H. G. Wells 1986. Towards a phonology of conversation: Turn-taking in Tyneside English. *Journal of Linguistics* 22. 411–437.
Mattingly, I. G. 1999. A short history of acoustic phonetics in the U.S. *Proc. 14th International Congress of Phonetic Sciences, San Francisco, USA*, 1–6.
Menzerath, P. & A. de Lacerda 1933. *Koartikulation, Steuerung und Lautabgrenzung. Eine experimentelle Untersuchung*. Berlin and Bonn: Dümmler.
Morais, J., L. Cary, J. Alegria & P. Bertelson 1979. Does awareness of speech as a sequence of phones arise spontaneously? *Cognition* 7. 323–331.
Mücke, D. & M. Grice 2014. The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation? *Journal of Phonetics* 44. 47–61.
Niebuhr, O. 2011. On the domain of auditory restoration in speech. In B. Kokinov, A. Karmiloff-Smith & N. J. Nersessian (eds), *European perspectives on cognitive science*, 1–6. Sofia: New Bulgarian University Press.
Niebuhr, O. & K. Kohler 2011. Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics* 39. 319–329.
Niebuhr, O. 2012. At the edge of intonation – The interplay of utterance-final F0 movements and voiceless fricative sounds. *Phonetica* 69. 7–27.
Niebuhr, O., K. Görs & E. Graupe 2013. Speech reduction, intensity, and F0 shape are cues to turn-taking. *Proceedings of14th Annual SigDial Meeting on Discourse and Dialogue*, 1–9. Metz, France.
Niebuhr, O. 2014. "A little more ironic" – Voice quality and segmental reduction differences between sarcastic and neutral utterances. *Proceedings of 7th International Conference of Speech Prosody*. 1–5. Dublin, Irland.
Niebuhr, O. & J. Hoekstra 2015. Pointed and plateau-shaped pitch accents in North Frisian. *Laboratory Phonology* 8. 1–35.
Niebuhr, O. 2016. Rich Reduction: Sound-segment residuals and the encoding of communicative functions along the hypo-hyper scale. *Proceedings of 7th Tutorial & Research Workshop on Experimental Linguistics*, 11–24. St. Petersburg, Russia.
Niebuhr, O. 2017. Clear speech, mere speech? How segmental and prosodic speech reduction shape the impression that speakers create on listeners. *Proceedings of 18th Interspeech Conference*, Stockholm, Sweden.
Nolan, F. 1992. The descriptive role of segments: Evidence from assimilation. In G. Docherty & D. R. Ladd (eds.), *Laboratory phonology*, 2. 261–280. Cambridge: Cambridge University Press.
O'Connor, M. 1983. Writing systems, native speaker analysis, and the earliest stages of Northwest Semitic orthography. In C. Myers & M. O'Connor (eds.), *The word of the Lord shall go forth*, 439–465. Winona Lake: Eisenbrauns.
Ogden, R. 2012. Prosodies in conversation. In O. Niebuhr (ed.), *Understanding Prosody: The Role of Context, Function and Communication*, 201–218. Berlin and New York: de Gruyter.
Ohala, J. J. 2004. Phonetics and phonology then, and then, and now. In H. Quene & V. van Heuven (eds.), *On speech and language: Studies for Sieb G. Nooteboom*, 133–140. Utrecht: LOT Occasional Series.
Origlia, A. & I. Alfano 2012. Prosomarker: a prosodic analysis tool based on optimal pitch stylization and automatic syllabification. *Proceedings of 8th International Conference on Language Resources and Evaluation (LREC)*, 997–1002. Istanbul, Turkey.

Plug, L. 2005. From words to actions: The phonetics of eigenlijk in two communicative contexts. *Phonetica* 62. 131–145.
Port, R. 2010. Rich memory and distributed phonology. *Language Sciences* 32. 43–55.
Sampson, G. 1985. *Writing systems*. Stanford: Stanford University Press.
Saussure 1959 [1916]. *Course in general linguistics*. New York: Philosophical library.
Schleicher 1861–1862. *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*. Weimar: H. Böhlau.
Schubotz, L., N. Oostdijk & M. Ernestus 2015. Y'know vs. you know: What phonetic reduction can tell us about pragmatic function. In S. Lestrade, P. de Swart & L. Hogeweg (eds.), *Addenda – Artikelen foor Ad Foolen*, 361–380. Nijmegen: Radboud University Press.
Simon, D., E. Burns & C. Virgo 2002. Old cases [Television series episode]. In D. Simon & R. Colesberry (prod.), *The wire*. New York: Home Box Office.
Siebert, H. 2003. *Der Kobra-Effekt. Wie man Irrwege der Wirtschaftspolitik vermeidet*. München: Piper.
Spitzer, L. 1921. *Italienische Kriegsgefangenenbriefe*. Bonn: Hanstein.
Sproat, R. & O. Fujimura 1993. Allophonic variation in English /1/ and its implications for phonetic implementation. *Journal of Phonetics* 21. 291–311.
Stampe, D. 1973. *A Dissertation on natural phonology*. PhD Diss. University of Chicago.
Strunk, J., F. Schiel & F. Seifart 2014. Untrained forced alignment of transcriptions and audio for language documentation corpora using WebMAUS. *Proceedings of 9th International Conference on Language Resources and Evaluation (LREC')*, 3940–3947. Paris, France.
Torreira, F. & M. Grice (in press). Melodic constructions in Spanish: Metrical structure determines the association properties of intonational tones. *Journal of the International Phonetic Association*.
Trubeckoj 1939. *Grundzüge der Phonologie*. Prague: Travaux du Cercle Linguistique de Prague.
Vennemann, T. 1972. On the theory of syllabic phonology. *Linguistische Berichte* 18. 1–18.
Vygotskij 1962 [1934]. *Thought and language*. Cambridge, MA: M.I.T. Press.
Wagner, P., A. Origlia, C, Avesani, G. Christodoulides, F. Cutugno, M. D'Imperio, M. D. Escudero, B. Gili Fivela, A. Lacheret, B. Ludusan, H. Moniz, A. Ní Chasaide, O. Niebuhr, L. Rousier-Vercruyssen, A. C. Simon, J. Simko, F. Tesser & M. Vainio 2015. Different parts of the same elephant: A roadmap to disentangle and connect different perspectives on prosodic promience. *Proceedings of 18th International Congress of Phonetic Sciences*, 1–5. Glasgow, UK.
Wesener, T. 1999. The phonetics of function words in German spontaneous speech. *AIPUK* 34. 327–377.
Wood, S. A. J. 1996. Assimilation or coarticulation? Evidence from the temporal co-ordination of tongue gestures for the palatalization of Bulgarian alveolar stops. *Journal of Phonetics* 24. 139–164.
Xu, Y. 2013. ProsodyPro – A tool for large-scale systematic prosody analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP)*, 7–10. Aix-en-Provence, France.