# 1 Getting started

> It is the holy grail of phonology to be sure of what the underlying form is.
>
> (William Labov, 21 May 2010, Manchester Phonology Meeting)

## 1.1 WHAT THIS BOOK IS ABOUT

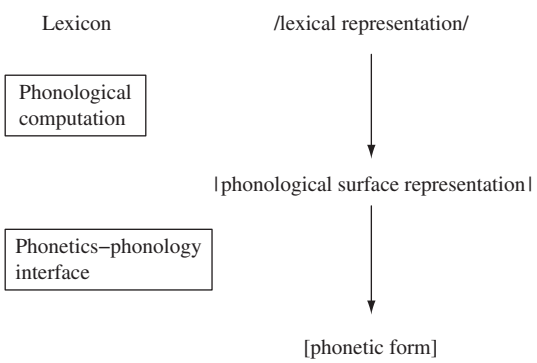
This book surveys the development of the concept of underlying representation or underlying form over the last 100 years or so within the field of generative phonology and its predecessors. We will consider phonological patterns and phenomena such as hypercorrection, linguistic experiments, statistical generalizations made over data corpora as well as theoretical arguments that have been used as arguments for underlying representations, their form and degree of (under) specification or the absence of such abstract entities. We will necessarily also look at the theoretical background that shaped our understanding of underlying representations at different times. Finally, a view of underlying representations will be converged on that sees two principles of economy as central in the determination of underlying representations: the avoidance of unnecessary information and the (over)generalization of alternation-inducing patterns to non-alternating forms as a strategy to achieve this goal of lexical parsimony. These principles, designed to maximize different aspects of lexical economy, actually stand in conflict with each other. It will be shown that these conflicts lead in some instances to underlying representations that defy common intuitions on economic lexical representations.

One cannot think about underlying representations without considering what their basic building blocks are. Accordingly, this book will also discuss this aspect of underlying representations. Originally, the smallest unit of phonology was thought to be the phoneme. As in physics, it turned out very soon that the atomic unit consisted of smaller elements or particles, the contrastive

features. The nature of these features and their contents are pragmatically defined as relating them to their articulation or to the corresponding acoustic signal, though we will see that there is considerable dissent when it comes to the details. On the basis of economy considerations, we will end up with a theory of features that connects the smallest building blocks of phonological representations with concepts from other domains of cognition, but especially with other modules of the language faculty.

Underlying representations are at the heart of modern phonological theorizing. Most contemporary, especially generative, phonological theories are mechanisms that map assumed more or less abstract underlying representations to much less abstract phonological representations, which are either regarded as instructions to the articulators or translated into such in a phonology–phonetics interface component.

(1) From the phonological lexicon to the phonetics[1]




While abstract phonemic representations were a mere theoretical construct in structuralist phonology, underlying representations received a more challenging status in generative phonology and even more so in psycholinguistic research. The criterion of psychological reality was imposed on the postulation/deduction of underlying representations such that present-day phonological research is trying to pin-point how the signifying part of atomic linguistic units is stored in the human mind.

The terms lexical representation and underlying representation are used in various ways in the literature. Some authors use the former term to refer to some intermediate representation that is created from the lexically stored form or underlying representation, while for other authors the underlying form is derived from a more abstract form. In this book I will try to avoid this terminological pitfall and use both terms, *underlying* and *lexical*, as referring to the state of phonological

units before any computation to prepare them for production has applied, i.e., the representation stored in the (mental) lexicon.

Reasoning about underlying representations in phonology has been guided and shaped to a considerable degree by two ideas. Ockham's razor, or the *lex parsimoniae*,[2] is a general principle determining theory formation not only in linguistics; it holds that if more than one explanation is available, the one making fewer assumptions has to be preferred. The other guiding idea is the belief that the human capacity for storage of memories is limited. To be able to store as much information as possible, this information has to be stored in the most economic way.

In our context this can be broken down into four dimensions of lexical economy.

(2) Lexical economy (Yip 1996: 766)

a. economy of individual lexical entries
b. economy of phoneme inventory
c. economy of phonotactic combinations
d. economy of paradigms

If individual lexical entries are stored in the most economic way, to leave room for as many of them as possible, they should contain as little information as necessary to be distinguished from one another. If the goal of mental storage is maximal accuracy, individual lexical entries should contain all information on every single rendition of every word or morpheme a language user has ever perceived or produced or even thought of and might even store linguistically irrelevant information about the context in which the realization happened, such as the weather, the hairstyle of the speaker, the shoes of an accidental passer-by in the situation, etc.

The same standard of information-saving measures holds for the economy of the phoneme inventory. According to Saussure, all that matters in language is contrast. Quite a few of the sounds we use in each language are not contrastive segments but predictable positional variants. In addition, not all phonetic/physical aspects of every contrastive segment are necessary to distinguish it from all other such segments. In the optimal case we can distil a small set of relatively abstract contrastive features for each language that are sufficient to account for the contrasts. If only these are stored in underlying representations, this also reduces the load of every lexical entry. According to Ockham's razor, the analysis with the smallest set of features wins. Extending this beyond the analysis of single languages one can hope to find the smallest set of features necessary

to describe all contrasts in all languages, with every language using a principled subset. Again, the simplicity metric prefers this theory to one that postulates different contrastive features for different languages.

In all languages we are aware of, sounds cannot be combined randomly inside words. They are (probably) organized in syllables and other prosodic units. Usually this kind of structure can be generated automatically and doesn't have to be memorized. Many languages allow only sequences of consonant–vowel–consonant–vowel . . . (CVCV . . .), rather than random combinations of Cs and Vs, in which case one could even go as far as to claim that a competent user of such a language doesn't even have to memorize whether a sound is a consonant or vowel since that is determined by its position in the string.

Finally, in most languages, words often consist of several morphemes. These morphemes can be recombined to form new words, just as words are constantly recombined to form new phrases, sentences and utterances. So if we are dealing with paradigms, it is more economic to just store the individual morphemes rather than every form in every paradigm. To take an example, an ordinary Italian verb has around 57 different forms, including participles and the infinitive, and the past participle alone is used in an additional 42 forms together with 42 different forms of auxiliary verbs. Simplicity and the observation of productivity or linguistic creativity suggest breaking these forms down into a verb stem and definitely less than 57 affixes. Since this army of verb forms is inflected for tense/mood/aspect as well as person and number we can break down every form into a stem or root and several affixes, which recur in the paradigm. If we assume that an average Italian knows roughly 8,000 verbs we reduce the number of lexical entries from (8,000 × 57 =) 456,000 to around 8,020.[3]

A question to ask: but why should our mind be so obsessed with economy when it comes to storage? We have plenty of memory space. The human brain contains an estimated 100 billion nerve cells.

Question 1: How economic are underlying forms?
Question 2: What do these forms consist of? Are there features?

If the answer to question 2 is yes, which features are these? What is in a feature? How many are there? Are they universal, i.e., already present at birth by genetic endowment, or learned? Are they language-specific?

One can add two more technically minded subquestions then: how do we arrive at underspecification (i.e., what is the best algorithm)? What are the boundaries of phonological data compression? Learnability might set a natural boundary in limiting the abstractness of such underlying representations, but also other factors, such as efficiency of computation and retrieval. After all, once a thought has turned up in our mind followed by the desire to dress this thought in language, we want to be fast and efficient in finding the necessary morphemes and words to express it and excessive abstractness might hamper this ambitious task. One can also assume that any listener wants to keep the job of interpreting a sound stream as efficient and fast as possible and the presence of more phonetic detail in lexical entries could help identify the right lexical entries.

The degree of abstractness, as well as the primitives of phonological representations in general and in particular of underlying representations, have been subject to a long debate starting with the introduction of underlying representations into modern linguistics by Bloomfield (1933). Before this, the Prague school, based on Saussure's work, introduced a revolutionary degree of abstractness with the archiphoneme (Jakobson 1939, Trubetzkoy [1939]1971).

While American structuralism focused on the identification of phonemes and their allophones based on the Bloomfieldian view of underlying form, generative phonology, especially from the publication of Chomsky & Halle's (1968) *The Sound Pattern of English* (*SPE*), concentrated more on phonological processes, i.e., the computational aspect of phonology. This led to a much more abstract view of underlying forms.

With the observation that on the one hand there are substantial restrictions on the form of underlying representations, i.e., the lexicon (morpheme structure rules; Halle 1959), and that on the other hand very often similar rules had to be assumed for the grammar mapping underlying forms to surface representations, the duplication problem arose (Kenstowicz & Kisseberth 1977).

While already early on in generative phonology predictable or redundant features were assumed to be filled in by feature filling or redundancy rules, a type of rule to be distinguished from actual phonological rules, the degree of abstractness, was furthered in the 1980s and early 1990s with the emergence of theories of underspecification, e.g., Radical Underspecification (Archangeli 1984, 1988, Archangeli & Pulleyblank 1989) and Contrastive Underspecification (Steriade 1987, 1995).

However, since the main objective of generative linguistics is to define the space of grammar, i.e., develop a theory of what is a possible natural language and what is not, the conclusion that underlying representations are highly abstract is not ineluctable. That underlying representations are much richer than had been concluded in the above-mentioned studies was argued for by Mohanan (1991) and Vaux (2003) among others.

In the 1990s, the introduction of Optimality Theory (OT; Prince & Smolensky 1993/2004) gave rise to a complete overhaul of the view of underlying representations. First of all, the basic idea behind the framework explicitly doesn't allow for restrictions on underlying representations, which solves the duplication problem via hypothesis. The question then is whether the theory has anything to say about underlying representations at all. While, in general, generative work assumes certain principles of economy applying in the determination of underlying forms (for example, every aspect of representation that is predictable is stripped off and taken care of by the grammar, such as syllable structure or redundant non-contrastive features), OT provides a fully automatic way to determine underlying representations: Lexicon Optimization (LO; Prince & Smolensky 1993/2004, Inkelas 1994, Itô, Mester & Padgett 1995 and others). Lexicon Optimization reverses the usual mechanism of candidate evaluation central to OT by evaluating competing underlying representations that are compatible with a given output form (i.e., that, if assumed as the input, would lead to an evaluation of this output candidate as the winner). The nature of the set of constraints, especially Faithfulness constraints, determines the underlying representation that is most like the corresponding output as the winning candidate.

Prince & Smolensky (1993/2004) already raise doubts on whether identity of underlying and surface forms is the desired outcome and entertain the option that other grammatical mechanisms besides candidate evaluation against the constraint hierarchy could be involved in the determination of underlying forms.

More recently, Lexicon Optimization has come under attack in two respects. On the one side it was doubted that Lexicon Optimization actually makes the predictions posited in earlier work, since the definitions of constraints as well as the theory of representation (Feature Geometry, *SPE*-style binary features, etc.) are actually crucial factors (see, e.g., McCarthy 2003; Krämer 2006a, b), as are productive patterns resulting in alternations (Harrison & Kaun 2000). A more radical criticism challenges the notion of LO altogether (Nevins & Vaux 2007). In a similar vein as the latter faction, McCarthy (2005)

introduces an OT-compatible reformulation of the free ride principle that leads to a higher degree of abstractness/underspecification in some underlying forms.

Some researchers in the realm of OT have abandoned the idea of underlying representations altogether and instead assume that the only representations grammar has access to are surface forms (e.g., Burzio 1996 *et seq.*). In such approaches correspondence relations are established between morphologically related output forms, mostly to explain phenomena within paradigms, such as paradigm levelling, paradigm uniformity, etc.

While all the former agree on the nature of (underlying) representations as composed of discrete categorical primitives (for an exception to this within OT, see, e.g., the work of Boersma), Exemplar Theory (e.g., Pierrehumbert 2001) departs from this common ground and regards words, morphemes or even individual phonemes as stored clouds of memories of actual realizations with all phonetic detail (linguistic or not), just like a box of high-resolution photographs (and just like with high-resolution photographs printed on low-quality paper or with cheap ink, the details fade away over time). This position is based on frequency effects found in the application of phonological processes. Roughly speaking, more frequent lexical items have been found to be more susceptible to neutralization than less frequent items. From this, advocates of Exemplar Theory conclude that the greater number of stored exemplars that come with higher frequency results in more variation in the stored signal and therefore higher flexibility in production, i.e., lower cost to neutralization.

Thus, generative phonology moves from underlying forms of extreme abstractness and extreme underspecification to underlying forms of a very low degree of abstractness, with full specification of all phonological features, to the rejection of underlying forms altogether and to a phonetically detailed view of underlying forms.

The great challenge in this field is that the nature and details of underlying representations cannot be observed directly, unlike surface patterns. We will see that the predictions of theories regarding underlying representations are far from unambiguous. One could likewise start with assumptions about underlying representations that then shape the theory, though priority has to be given to indirect observation through data collection and experimentation. However, there is often more than one conclusion to be drawn from naturally occurring linguistic patterns, statistical observations and the like, as has also been indicated in this short summary of the ongoing discussion on underlying representations in phonology.

## 1.2 WHERE YOU FIND WHAT

The book covers the following ground. Chapter 2 goes back to the early twentieth century, discussing Saussure and Trubetzkoy in particular and their ideas of contrast and phonemes as well as the relevance of categorical contrast in phonology. We will also look at Trubetzkoy's methodology to diagnose contrastive segments and touch on his ideas about contrastive segments as a system. I hope to elucidate to the reader the foundations of modern theorizing and research on underlying representations in phonology in this way.

Chapter 3 is concerned with the abstractness of underlying representations. This increased as a by-product of the development of a more sophisticated theory of transformation from the representation assumed to be phonologically relevant to the form that is passed on to the phonetics, which consists of non-categorical physical events. Saussure and Trubetzkoy prepared the ground for what was later termed underspecification and Jakobson's and Halle's work moved further in this direction. Generative phonology of the 1950s was based on these scholars' work until the idea of underspecification of redundant or predictable segmental information was challenged and by and large abandoned for at least a decade.

Chapter 4 reviews the arguments to reconsider theories of underspecification and how these developed from the 1980s on. In this chapter we also have a look at more systematic ways to derive different degrees of underspecification, before we turn to an argument against systematic underspecification and the use of blanks (underspecification) in analyses of lexical exceptions.

Chapter 5 is dedicated to the opposite extreme view. Usage-based phonology, or Exemplar Theory, holds that language computation runs statistics over huge corpora of experienced and memorized linguistic data and that linguistic patterns and language change can be explained this way. At the end of this chapter we try to reconcile these opposing views.

Chapter 6 reviews studies, mostly from the 1990s and the first decade of the twenty-first century, that provide insight into whether the human brain operates with abstract categorical units and whether phonological features are ever underspecified.

The content of phonological features is the subject of Chapter 7. In Chapters 2–6 standard phonological features are used, which are defined mostly in articulatory and partially acoustic terms. Chapter 7 looks at the motivation of feature definitions and concludes that there

is an alternative way of defining phonological features on a broader cognitive basis.

While *The Sound Pattern of English* (*SPE*) and its successor, *Lexical Phonology*, provide the background for most of the discussion in Chapters 3 and 4, Chapter 8 is dedicated to the issue of underlying representations in Optimality Theory. The chapter starts with a sketch of the basics of the theory and then discusses its mechanism of Lexicon Optimization (LO). LO will be shown not only to make false predictions (compared with some of the results of Chapter 6) but also to be based on a misunderstanding. The last part of the chapter develops a revised version of LO that generates underlying forms more in line with Chapter 6.

Finally, in Chapter 9 we will look back to see if we have learned anything from the preceding 210 pages.

## 1.3 WHAT YOU WON'T FIND HERE

There are quite a few aspects of phonological research that one could imagine including in a book with this title that you won't find in here. The issue of feature values is discussed in many places in the book, though a discussion of whether contrastive features should be binary, privative or multivalued is not really undertaken.

The internal structure of segments is not discussed either. If, for example, the trees of Feature Geometry are universal we (by simplicity) only need features, which we then can organize into the universal tree when needed in the computation to explain feature interaction and class behaviour. If some aspects of phonological contrast can be represented in the tree, as indicated (but not discussed) in Chapter 7, then this bears on the quest for the nature and form of underlying representations.

We will merely touch upon whether prosodic structure, such as syllables, syllable constituents, stress and feet and the like, has to be part of the mental representation of morphemes. Section 4.5, where this is taken up, only brushes over underlying syllable structure in a very cursory way, though there is extensive literature on the representation of lexical stress and foot structure.

The nature of tones and tonal contrasts is not covered at all. The lexical properties that underlie allomorphy and allomorph selection are potentially illuminating pieces of the puzzle and would have deserved a separate chapter if not a separate volume.

Finally, speech errors, language acquisition and language change potentially provide insight into the nature of the mental representation of the phonological side of linguistic units. These sources of evidence are referred to in many places in the book, but not discussed separately or thoroughly.

Thus, there is plenty of material left for a second volume.

## 1.4 HOW TO USE THIS BOOK

The description of the series this book is published in suggests that one of the ways this book is going to be used is as a textbook for advanced undergraduate or graduate students. In Norway, or at least at the University of Tromsø, where I have been teaching for the last six years, a course usually consists of thirteen weeks of lectures, interrupted by the occasional reading week. So, why doesn't this book have thirteen chapters then? I personally like the number thirteen, since it is a prime number, and it would not have been any problem to write another four chapters about different aspects of underlying representations. However, prime numbers are difficult to divide into teaching blocks of equal weight. Besides, at most universities, courses last for about ten to twelve weeks. There is another, more important issue with numbers here. Whenever I use a textbook in one of my courses that has eleven or twelve chapters, we either don't make it through all the chapters (and, remember, I have thirteen weeks to get through them), because we need more time for one or several topics, or I have to cut one or two chapters because there are some issues that are not covered in the book and I think they should have been. In the optimal case you won't have either reason to skip a chapter of this book, since there are only nine of them. The reading suggestions at the end of each chapter are of two types, they can either be used to cover the topic of the chapter or an aspect of it in more depth or to go beyond the scope of the chapter and explore a related topic. In this way every unit can be expanded and dwelled on for more than one session.

Since the topic of the book is rather broad, though it actually only covers a subset of the theoretical issues and relevant empirical areas, there is ample room to add topics according to the lecturer's and the students' interests. The bearing of different types of allomorphy or of prosodic morphology on our understanding of underlying representations, for example, is not touched on here at all. The question of whether syllable structure is ever stored lexically is only dealt with briefly (in Section 4.5 and by the prediction of the absence of syllable