



Copying (ii): Coarticulation

e.g. keep [k̥]  
 cool [k̠]  
 cart [k̠]

Rules:  $k \rightarrow k̥ / \_ \text{---} \begin{matrix} V \\ [-back] \end{matrix}$   
 $k \rightarrow k̠ / \_ \text{---} \begin{matrix} V \\ [+rɹnd] \end{matrix}$   
 $k \rightarrow \underline{k} / \_ \text{---} \begin{matrix} V \\ [+back] \end{matrix}$

+ denotes advanced articulation  
 ̠ denotes lip-rounding  
 ̡ denotes retracted articulation

Insertion: Epenthesis

e.g. mince [mɪnts]  
 pence [pɛnts]

Rule: ns → nts

Deletion: Elision

e.g. sandwich [sænɪtʃ]

Rule: nd → n

Permutation: Metathesis

e.g. burnt [brʌnt]

Rule: ur → ru

The problems inherent in this approach are many:

1. Deletion rules can make Context-Sensitive grammars undecidable. (Salomaa 1973:83, Levelt 1976:243, Lapointe 1977:228, Berwick and Weinberg 1984:127)
2. Non-monotonicity makes for computational complexity.
3. There is no principled<sup>1</sup> way of limiting the domain of rule application to specific linguistic domains, such as syllables.
4. Using *sequences* as data-structures is really only plausible if all speech parameters change with more-or-less equal regularity.

<sup>1</sup>N.B. The use of labelled brackets to delimit domains is a completely unrestricted mechanism for partitioning strings.

e.g. "chip"

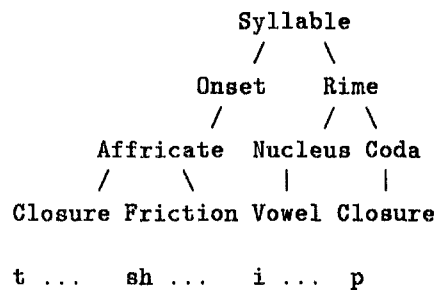


Figure 1: Richer structure in phonological representations

In partial recognition of some of these problems, phonologists have been attempting to reconstruct the transformational component as the epiphenomenal result of several interacting general "constraints". Numerous such "constraints" and "principles" have been proposed, such as the Well-Formedness Condition (Goldsmith 1976 and several subsequent formulations), the Obligatory Contour Principle (Leben 1973), Cyclicity (Kaisse and Shaw 1985, Kiparsky 1985), Structure-Preservation (Kiparsky 1985), the Elsewhere Condition (Kiparsky 1973) etc. While this line of research is in some respects conceptually cleaner than primitive transformational grammars, there has been no demonstration that a "principle"-based phonology is indeed more restrictive than primitive transformational phonology in any computationally relevant dimension.

**A declarative model of speech** For the last few years, I have been developing a "synthesis-by-rule" program which does not employ such string-to-string transformations (Coleman and Local 1987 forthcoming; Local 1989 forthcoming; Coleman 1989).

The basic hypothesis of this (and related) research is that there is a trade-off between the richness of the rule component and the richness of the representations (Anderson 1985). According to this hypothesis, the reason why transformational phonology needs to use transformations is because its data structure, strings, is too simple. Consequently, it ought to be possible to considerably simplify or even completely eliminate the transformational rule component by using more elaborate data structures than just well-ordered sequences of letters or feature-vectors. For instance if we use graphs (fig. 1) to represent phonological objects, then instead of copying, we can im-

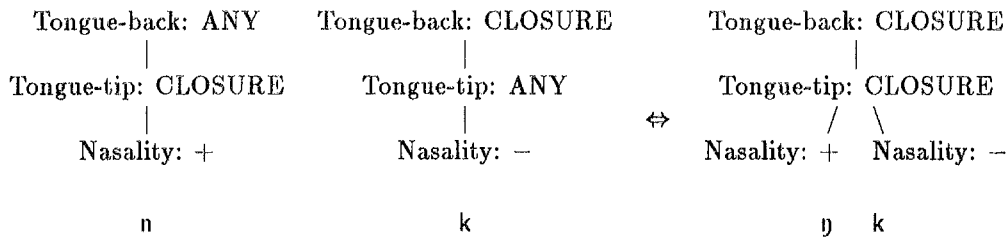
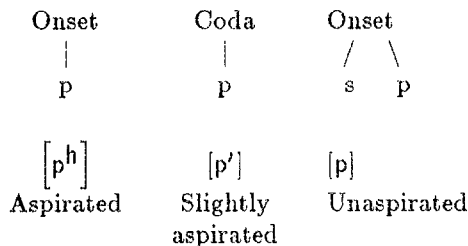


Figure 2: Declarative characterisation of assimilation

plement harmony phenomena using the structure-sharing technique.

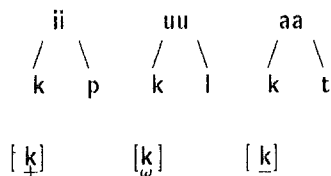
Incorporating richer data-structures allows many if not all rewriting rules to be abandoned, to the extent that the transformational rewrite-rule mechanism can be ditched, along with the problems it brings. Consider how the “processes” discussed above can be given a declarative (or “configurational”) analysis.

*Allophony* can be regarded as the different interpretation of the same element in different structural contexts, rather than as involving several slightly different phonological objects instantiating each phoneme.



*Assimilation* can also be modelled non-destructively by unification (fig. 2).

*Coarticulation* is simple to model if parametric phonetic representations may be glued together in parallel, rather than simply concatenated. Consonants may then overlaid over vowels, rather than simply concatenated to them (Öhman 1966, Perkell 1969, Gay 1977, Mattingly 1981, Fowler 1983). If required, this analysis can also be implemented in the phonological component, using graphs of the ‘overlap’ relation (Griffen 1985, Bird and Klein 1990), e.g.:



It is now common to analyse *epenthesis*, not as the insertion of a segment into a string, but as due to

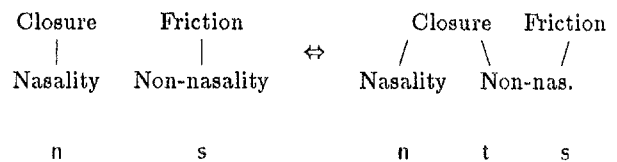


Figure 3: Declarative characterisation of epenthesis

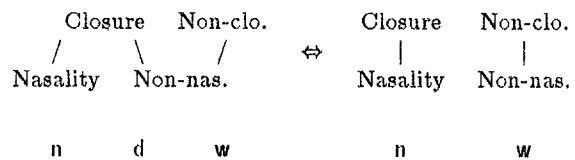


Figure 4: Declarative characterisation of elision

minor variations in the temporal coordination of independent parameters (Jespersen 1933:54, Anderson 1976, Mohanan 1986, Browman and Goldstein 1986) (fig. 3).

It has been demonstrated (Fourakis 1980, Kelly and Local 1989) that epenthetic elements are not phonetically identical to similar non-epenthetic elements. The transformational analysis, however, holds that the phonetic implementation of a segment is dependent on its features, not its derivational history (“a [t] is a [t] is a [t]”), and thus incorrectly predicts that an epenthetic [t] should be phonetically identical to any other [t].

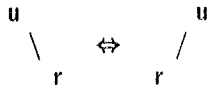
*Elision* is the inverse of epenthesis, and is thus in some sense “the same” phenomenon, taking the “un-elided” form as more primitive than the “elided” form, a decision which is entirely meaningless in the declarative account (fig. 4)

*Metathesis* is another instance of “the same” phenomenon i.e. different temporal synchronisation of an invariant set of elements. Epenthesis, Elision and Metathesis may all be regarded as instances of the more general phenomenon of non-significant variability in the timing of parallel events.

Figure 5: Phrase structure grammar of English phoneme strings

$Word$ [+inflected]	$\rightarrow$	$Word$ [-inflected]	$Inflection$	Inflection e.g. cat+s
$Word$ [-inflected]	$\rightarrow$	$Word$ [-inflected]	$Word$ [-inflected]	Compounding e.g. black+bird
$Word$ [-Latinate]	$\rightarrow$	$Prefix^*$ [-Latinate]	$Word$ [-inflected]	$Suffix^*$ [-Latinate]
$Word$ [+Latinate]	$\rightarrow$	$Stress$ [+Latinate]	$\circ$	$Morphology$ [+Latinate]
$\circ$ denotes complete constituent overlap				
$Stress$ [+Latinate]	$\rightarrow$	$Non\_final\_feet$	$Foot$	
$Non\_final\_feet$	$\rightarrow$	$Foot$ [+initial]	$Foot^*$	
$Foot$	$\rightarrow$	$Syllable$ [+stress] [+heavy]	$\left( \begin{array}{c} Syllable \\ [-stress] \\ [-heavy] \end{array} \right)$	
$Foot$	$\rightarrow$	$Syllable$ [+stress]	$\left[ \begin{array}{c} Syllable \\ [-stress] \\ [-heavy] \end{array} \right]$	$\left[ \begin{array}{c} Syllable \\ [-stress] \\ [-heavy] \end{array} \right]$
$Morphology$ [+Latinate]	$\rightarrow$	$Prefix^*$ [+Latinate]	$Stem$ [+Latinate]	$Suffix^*$ [+Latinate]
$Syllable$ [αheavy]	$\rightarrow$	$(Onset)$	$Rime$ [αheavy]	
$Onset$ [αvoi]	$\rightarrow$	$Affricate$ [αvoi]		
$Onset$ [-voi]	$\rightarrow$	$Aspirate$		
$Onset$ [αvoi]	$\rightarrow$	$\left( \begin{array}{c} Obstruence \\ [αvoi] \end{array} \right)$	$(Glide)$	
$Obstruence$ [-voi]	$\rightarrow$	$([s]), Closure$		(Either order)
Constraint: in onsets, [s] < Closure				
$Rime$ [αheavy]	$\rightarrow$	$Nucleus$ [αheavy]	$\left( \begin{array}{c} Coda \\ [αheavy] \end{array} \right)$	
$Nucleus$	$\rightarrow$	$Peak$	$Offglide$	

etc. etc.



As well as these relatively low-level phonological phenomena, work in Metrical Phonology (Church 1985) and Dependency Phonology (Anderson and Jones 1974) has shown how stress assignment, a paradigm example for transformational phonology, can be given a declarative analysis.

### Overview of text-to-parameter conversion in the YorkTalk system

1. Each symbol in the input text string is translated into a column-vector of distinctive phonetic features (nasal, vowel, tongue-back, etc.) Sequences of letters are thus translated into sequences of feature-structures.
2. The sequence of feature-structures is parsed. This process translates the sequence into a directed graph representing the phonological constituent structure of the utterance.
3. The phonological structure is traversed and an *interpretation function* applied at each node to derive a phonetic parameter matrix.

Parsing is done using a Phrase Structure Grammar of phoneme strings. A very simplified version of such a grammar is fig. 5. I have implemented several such grammars so far, including a DCG implementation and a PATR-II-like implementation. With one or two simple extensions to the grammar formalism, it is also possible to parse re-entrant (e.g. ambisyllabic) structures and other overlapping structures, such as those arising from bracketting "paradoxes". The resulting graphs are thus not trees, but directed acyclic graphs.

In computational syntactic theory, one of the main uses for the parse-tree of a string is to direct the construction of a compositional (Fregean) semantic interpretation, according to the rule-to-rule hypothesis (Bach 1976). In the YorkTalk system, the same approach is employed to assign a phonetic interpretation to the phonological representation. A second, theory-internal motivation for constructing rich parse-graphs of the phonemic string is that it enables the phoneme string to be discarded completely, thus liberating the phonetic interpretation function from the sequentiality and other undesirable properties of segmental strings.

After the phonological graph has been constructed by the parser, a head-first graph-traversal algorithm maps the (partial) phonological category of each node into equations describing the time-dependent

motion of the synthesis parameters for specified intervals of time. These parametric time-functions are finally instantiated with actual numbers representing times, in order to derive a complete matrix of (parameter, value) pairs.

As well as being computationally "clean", this method of synthesis has the additional merit of being genuinely non-segmental in (at least) two respects: there are no segments in the phonological representations, and there is no cross-parametric segmentation in the phonetic representations. The resulting speech does not manifest the discontinuities and rapid cross-parametric changes which often cause clicks, pops, and the other disfluencies which typify some synthetic speech. On the contrary, the speech is fluent, articulate and very human-like. When the model is wrong in some respect, it sounds like a speaker of a different language or dialect, or someone with dysfluent speech. For all these reasons, the YorkTalk model is attracting considerable interest in the speech technology industry and research community, a circumstance which I hope will promote a widespread change of approach to computational phonology in future.

### References

- [1] Allen, J., S. Hunnicutt and D. Klatt. 1987. *From Text to Speech: The MITalk System*. Cambridge University Press.
- [2] Anderson, J. and C. Jones. 1974. Three theses concerning phonological representations. *Journal of Linguistics* 10, 1-26.
- [3] Anderson, S. R. 1976. Nasal Consonants and the Internal Structure of Segments. *Language* 52.2 326-344.
- [4] Anderson, S. R. 1985. *Phonology in the Twentieth Century*. University of Chicago Press.
- [5] Bach, E. 1976. An extension of classical transformational grammar. *Problems in Linguistic Metatheory, Proceedings of the 1976 Conference at Michigan State University*, 183-224.
- [6] Berwick, R. C. and A. S. Weinberg. 1984. *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition*. Cambridge, Massachusetts: M. I. T. Press.
- [7] Bird, S. and E. Klein. 1990. Phonological Events. To appear in *Journal of Linguistics* 26 (1).
- [8] Browman, C. P. and L. Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3, 219-252.
- [9] Chomsky, N. 1963. Formal Properties of Grammars. In R. D. Luce, R. R. Bush and

- E. Galanter, eds. *Handbook of Mathematical Psychology* Vol. II. New York: John Wiley.
- [10] Church, K. 1985. Stress Assignment in Letter to Sound Rules for Speech Synthesis. In *23rd Annual Meeting of the Association for Computational Linguistics* Proceedings.
- [11] Coleman, J. S. and J. K. Local. 1987 forthcoming. Monostratal Phonology and Speech Synthesis. To appear in C. C. Mock and M. Davies (eds.) In press. *Studies in Systemic Phonology* London: Francis Pinter.
- [12] Coleman, J. S. 1989. The Phonetic Interpretation of Headed Phonological Structures Containing Overlapping Constituents. ms. (Currently submitted to *Phonology*)
- [13] Fourakis, M. S. 1980. A Phonetic Study of Sonorant Fricative Clusters in Two Dialects of English. *Research in Phonetics* 1, Department of Linguistics, Indiana University.
- [14] Fowler, C. A. 1983. Converging Sources of Evidence on Spoken and Perceived Rhythms of Speech: Cyclic Production of Vowels in Monosyllabic Stress Feet. *Journal of Experimental Psychology: General* Vol. 112, No. 3, 386-412.
- [15] Gay, T. 1977. Articulatory Movements in VCV Sequences. *Journal of the Acoustical Society of America* 62, 182-193.
- [16] Gazdar, G. 1987. COMMIT ==> \* PATR II. In *TINLAP 3: Theoretical Issues in Natural Language Processing 3. Position Papers*. 39-41. Association for Computational Linguistics.
- [17] Goldsmith, J. 1976. *Autosegmental Phonology*. Indiana University Linguistics Club.
- [18] Griffen, T. D. 1985. *Aspects of Dynamic Phonology* Amsterdam studies in the theory and history of linguistic science. Series 4: Current issues in linguistic theory, vol. 37: Benjamins.
- [19] Hertz, S. R. 1981. SRS text-to-phoneme rules: a three-level rule strategy. *Proceedings of ICASSP 81*, 102-105.
- [20] Hertz, S. R. 1982. From text to speech with SRS. *Journal of the Acoustical Society of America* 72(4), 1155-1170.
- [21] Hertz, S. R., Kadin, J. and Karplus, K. 1985. The Delta rule development system for speech synthesis from text. *Proceedings of the IEEE* 73(11), 1589-1601.
- [22] Hertz, S. R. forthcoming. The Delta programming language: an integrated approach to non-linear phonology, phonetics and speech synthesis. In J. Kingston and M. Beckman, eds. *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge University Press.
- [23] Jespersen, O. 1933. *Essentials of English Grammar*. London: George Allen and Unwin.
- [24] Johnson, C. D. 1972. *Formal Aspects of Phonological Description*, Mouton.
- [25] Kaisse, E. and P. Shaw. 1985. On the Theory of Lexical Phonology. *Phonology Yearbook* 2, 1-30.
- [26] Kelly, J. and J. K. Local. 1989. *Doing Phonology*. Manchester University Press.
- [27] Kiparsky, P. 1973. "Elsewhere" in *Phonology*. Indiana University Linguistics Club.
- [28] Kiparsky, P. 1985. Some Consequences of Lexical Phonology. *Phonology Yearbook* 2, 82-136.
- [29] Lapointe, S. 1977. Recursiveness and deletion. *Linguistic Analysis* 3: 227-265.
- [30] Leben, W. 1973. *Suprasegmental Phonology*. Ph.D. dissertation, M. I. T.
- [31] Levelt, W. J. M. 1976. Formal grammars and the natural language user: a review. In A. Marzollo, ed. *Topics in Artificial Intelligence* CISM courses and lecture notes no. 256. Springer.
- [32] Local, J. K. 1989. Modelling assimilation in non-segmental rule-free synthesis. To appear in D. R. Ladd and G. Docherty, eds. *Papers in Laboratory Phonology II* Cambridge University Press.
- [33] Mattingly, I. G. 1981. Phonetic Representations and Speech Synthesis by Rule. In T. Myers, J. Laver and J. Anderson, eds. *The Cognitive Representation of Speech*. North-Holland.
- [34] Mohanan, K. P. 1986. *The Theory of Lexical Phonology*. D. Reidel.
- [35] Öhman, S. E. G. 1966. Coarticulation in VCV Utterances: Spectrographic Measurements. *Journal of the Acoustical Society of America* 39, 151-168.
- [36] Perkell, J. S. 1969. *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study* Cambridge, Massachusetts: M. I. T. Press.
- [37] Salomaa, A. 1973. *Formal Languages*. New York: Academic Press.
- [38] Van Leeuwen, H. C. 1987. Complementations introduced in linguistic rewrite rules. *Proceedings of the European Conference on Speech Technology 1987* Vol. 1, 292-295.
- [39] Van Leeuwen, H. C. 1989. A development tool for linguistic rules. *Computer Speech and Language* 3, 83-104.