Tobias Scheer  
Abstract Mfm 17, special session on the history of phonology  

Modularity and translation in structuralist and generative phonology  

The modular perspective on the architecture and the functioning of the mind/brain is one of the deepest layers of generative thinking: Chomskyan linguistics are the application of the cognitive science programme that was laid out in the 50s to a specific area of cognitive activity. Modularity is the idea that cognitive computation is specialized, rather than all-purpose: the mind/brain is made of a number of computational units that can only carry out a very narrowly defined task, which are insensitive to eventual teleological goals of the global system, and which work only with a specific vocabulary (domain specificity). Of course modules network, and a necessary condition of intermodular communication that follows from domain specificity is translation: a given module can understand only its own vocabulary.  

The talk aims at providing an overview of modularity and associated translation in structuralist and generative phonology regarding two specific modules, morpho-syntax and phonology (whether morphology and syntax represent one or two modules is orthogonal to the discussion). The conception of morpho-syntax and phonology as two ontologically distinct computational units is shared by the structuralist and the generative approach to language. In the former, so-called Level Independence organises grammatical activity: the use of morpho-syntactic information in phonology is prohibited. However, this hardwired prohibition is circumvented by juncture phonemes, which is the form that translation takes in structuralist phonology: morpho-syntactic information is converted into phonological vocabulary (a phoneme).  

The basic phonological vocabulary has changed over time, but translation has remained constant: after juncture phonemes, its output were segments in SPE (the hash mark #, which was a [-segment] segment since segments were the basic units now) and autosegmental trees (the Prosodic Hierarchy) when phonology was autosegmentalised in the early 80s.  

In early generative phonology, phonological computation was allowed to take untranslated morpho-syntactic information into account: while morpho-syntactic structure was translated and appeared as (clusters of) hash marks in phonology, the input to phonological computation was a bracketed and labelled string where morpho-syntactic labels (N, V, A) could be directly accessed by rules. After a conflict with so-called direct syntax approaches that upheld this option, emerging Prosodic Phonology imposed Indirect Reference in the mid-80s, a position that was largely accepted by the field. Indirect Reference installs true modularity by outlawing any reference to untranslated morpho-syntactic information in phonology (curiously enough, though, this was done without any reference to modularity).  

Another debate of the 80s could also have been decided by modularity, but was led without reference to it: interactionism (i.e. the interleaving of concatenation of pieces and phonological interpretation) was introduced by Lexical Phonology, but vigorously rejected by orthodox generative phonology and namely by Morris Halle, who upheld that all concatenation must precede all interpretation. It is obvious from hindsight that cyclic derivation (i.e. interpretation of pieces following morpho-syntactic structure, i.e. from the most to the least embedded item), a key property of generative linguistics, is only compatible with modularity if the communication with phonology is interactionist. This is precisely the evolution that current syntactic theory has taken, and which appears to be consensual in generative quarters today (including Morris Halle, who is actively engaged in interactionist Distributed Morphology): derivation by phase is the application of the Lexical Phonology insight to syntax and by syntacticians. It is the spine of current minimalist syntax and within this frame has acquired a new quality: derivation by phase is the instrument that unburdens workbench memory by cutting the derivation of a whole sentence into pieces. Shaping grammatical theory according to extra-grammatical conditions of cognitive/brainal implementation is indeed the motor of minimalist thinking.  

Finally, current interface practice in OT is critically reviewed in the light of modularity and translation: we face a situation where Indirect Reference and the request for translation is commonly ignored and violated without discussion or comment. A salient aspect of the behaviour of (classical) OT is anti-cyclicity, i.e. the rejection of cyclic derivation because of its serial character. OT has
produced a whole anti-cyclicity literature (which has an important intersection with the anti-opacity literature) that proposes alternative, strictly parallel ways of communicating with morpho-syntax: co-phonologies, indexed constraints, OO faithfulness and so-called interface constraints. On the other hand, DOT and Stratal OT uphold the derivational/interactionist architecture of Lexical Phonology. For some reason, classical OT has extended anti-derivationalism from phonological computation (i.e. events within a module) to the relationship among modules. This appears to be a category mistake: nothing withstands derivational intermodular communication (which is how modularity works elsewhere in the cognitive system) combined with strictly parallel computation inside the phonological module.

Diagnostics for OT's misty and largely unreflected relationship with modularity are the following: 1) customary and uncontradicted violations of Indirect Reference: ALIGN and WRAP constraints make constant reference to morpho-syntactic structure and labels; interface constraints such as FAITH-root and FAITH-affix make reference to designated morpho-syntactic categories (even reference to individual morphemes is not a problem: this is the reincarnation of SPE-practice where rules were supplemented with morphological diacritics); 2) mapping (of morpho-syntactic into phonological prosodic categories) is done in the phonological constraint hierarchy (rather than outside of the phonology as was the case in Prosodic Phonology): ALIGN and WRAP are interspersed with purely phonological constraints; 3) constraints whose formulation combines phonological and morphological instructions are commonplace.

All this is incompatible with a modular view of the cognitive system in general, and of grammar in particular. A reason for what I call the scrambling tropism of OT (i.e. the tendency to have everything computed in the same constraint chamber and thereby to eliminate modular contours, also with phonetics, and also incorporating more and more aspects of the lexicon) may be its empiricist origins: the self-understanding of OT is certainly generative, but this is only half of its genetic endowment. The central idea of OT, parallel computation, was developed in the second half of the 80s by connectionism, the empiricist challenger of the standard modular and rationalist theory of the cognitive system. Connectionism promotes PDP, Parallel Distributed Processing, and the question is whether it is possible/reasonable to cherry-pick just the two Ps (parallel processing) while refusing the D. Distributed processing means that computation is all-purpose, rather than specialized, i.e. the exact opposite of the modular idea. Prince & Smolensky (1993, chapter 10.2) are explicit on the fact that OT does cherry-picking in the connectionist toolbox – the question is whether this option is viable: the systematic violation of modularity in current OT practice is probably not unrelated to the anti-modular conception of the mind/brain that lies at the heart of connectionism.

The goal of the talk is not merely historiographic: the constancy of translation and its various incarnations over time raise the question what translation should look like. One property of translation that is shared by all structuralist and generative incarnations is that its output has only been thought of in terms of diacritics: of course juncture phonemes /#/ and /+/ are not real phonemes like /a/, /p/ etc., and of course #s are not segments like [a,p] etc. (something that was demonstrated early on by Pyle 1972), and of course the Prosodic Word, \( \omega \), is not anything that is known in phonology. Common to these items is their absence in phonological processes that are unconditioned by extra-phonological information (such as, say, a palatalisation): they only enter the scene when extra-phonological information plays a role, and their exclusive function is to carry this information (they act like a buffer that stores and unstores information). They are thus diacritics, i.e. phonological aliens. If phonology is a module, however, it is unable to understand alien vocabulary. In a sound interface theory, the output of translation must therefore not be diacritic: candidate carriers of morpho-syntactic information reduce to representational items that pre-exist in phonology and are also used for purposes other than just the transmission of boundary information, i.e. in phonological processes that are not conditioned by any extra-phonological information.

I call this requirement for using only truly phonological vocabulary as the output of translation Direct Interface: there must be no diacritic mediation between phonology and morpho-syntax.