

On overgeneration

Laura Bafile
University of Ferrara

From the beginning of Element Theory up to present day, successive revisions, almost without exceptions, have eliminated elements from the original inventory, in order to make it smaller. This development is certainly a result of the effort to give an adequate representation of relevant phonological categories, and to keep them apart from phonetic detail. However, the argument generally adduced in support of inventory reduction is, almost without exceptions, the prevention of overgeneration. Notably, this concern attracts considerable attention in ET work, and it is not by chance that almost all the proponents of a reduction of the original ten-element set (Harris and Lindsey 1995) mention the danger of overgeneration. (e.g. Backley 1993: 301, Jensen 1994: 72-73, Charette and Göksel 1996: 2, Backley 2011: 142).

This paper would like to propose a different point of view, whereby overgeneration is no relevant issue for a model that aims at contributing to a theory of grammar, and should not be a concern in establishing the melodic primes.

This proposal is based on three kinds of arguments:

1. The reduction of the element set, as an answer to the danger of overgeneration, may raise the opposite risk of undergeneration. Undergeneration is to be intended in the twofold sense of insufficient capacity of expressing phonemic contrasts of languages and, more interestingly, insufficient capacity of capturing relevant generalisations concerning the phonological behaviour of segments, according to the *phonological epistemological principle* (Kaye 2005). An example of this issue is the unification of velarity and labiality in a single prime (cf. Backley 2011), that expresses the fact that velar and labial consonants and rounded vowels make a natural class, on the one hand, but on the other hand fails to represent the fact that labiality and velarity may be independent properties (see Scheer 1999 for discussion).

2. The risk of undergeneration (especially in the first sense mentioned above) has encouraged the quest of alternative strategies in order to grant the expressive potential to a restricted set of element (e.g. the *licensing constraints* approach, Kaye 2001, Charette and Göksel 1996). These strategies themselves, however, may give birth to unrestricted models and pose problems of arbitrariness that should be taken into account as much as overgeneration is. Headedness is a major question in this regard. The traditional conception of headedness in dependency phonology is a relational one, whereby the head is the prominent element in a segmental expression; prominence may be conceived both in the sense of contributing the stronger acoustic property of the segment and in the sense of affecting its phonological behaviour more strongly. From this point of view, the approaches that allow double-headed expressions and, even more, those that admit that a single-element expression may be headed or headedless, thus corresponding to two different segments, crucially weaken of the relational conception of dependency (cf. Backley and Nasukawa 2009, Backley 2011, 2012). There is no doubt that the theorisation about a looser conception of headedness has been significantly motivated by the need of preventing effects of undergeneration, as this passage by Backley and Nasukawa (2009: 54) shows: «This headship relation serves two important purposes: first, it increases the number of possible melodic expressions – and thus, the number of contrasts – that the model can generate [...]».

3. Overgeneration refers to the capacity of the model to generate, by means of element combination, a set of segments much larger than the segmental inventories that can be observed in natural languages, and also larger than the inventory of contrastive sounds globally used by the world's languages. The potentially dramatic consequences of this perspective on phonological theory are expressed by Pöchtrager (2006: 14): «P[honological] E[xpression]s only encode what is phonologically relevant, and current estimates are that the number of expressions needed will be well below 100. Any theory generating more than that is certainly wrong». However, a less drastic conclusion is available. Typological studies indicate that most languages have segmental

inventories containing from 20 to 40 units, although, for example, in *UCLA Phonological Segment Inventory Database* (UPSID), the minimal cardinality is 11 (in two languages) and the maximal is 141 (in one language). Besides, in UPSID there are 919 different segments (although many of them would probably prove to be just phonetic variants of the same phonological categories) but the 80% of them only appear in ten or fewer languages, which means that almost all the languages of the sample make up their inventory by taking segments from a subset of 184 units. Thus, on the one hand, languages show strong preference for a subset of all attested cardinalities and for a subset of all attested segments. On the other hand, within the range of preferred subsets, languages vary to a large degree.

We can hypothesise that this state of affairs depends on general factors, not specifically linguistic. Properties of perception and articulation obviously limit the number of sounds that can be pronounced and discriminated by human sensory-motor system, i.e. the set of all possible segments. As for the number of segments that each language uses, we can reasonably conjecture that the lower limit may be due to the fact that a smaller number of segments would be not sufficient to the expression of lexical contrasts; the upper limit may depend on requirements of optimal computation, not uniquely linguistic in nature. Non-specifically linguistic considerations can also explain the preferred ranges observable in most languages, concerning both the dimension of inventories and the quality of segments. To sum up, if a part of the logically possible combinations of elements is excluded by non-linguistic factors, it follows that grammar is not what determines how many and which segments languages may have, and thus, that overgeneration is not an issue in a theory of grammar.

As Reiss (2012: 189) observes, overgeneration is not problematic if we characterise sets of features intensionally, i.e. from the perspective of language acquisition and of explanatory adequacy. From the viewpoint of language acquisition, this means that there is no reason to hypothesise that language learners start from the assumption that each one of the elements must combine with each other; and even less, that learners pursue the goal of evaluating the thousands of logically possible segmental inventories in order to single out the one that matches the target inventory.

In this perspective, overgeneration has no relevance to a theory of internal/intensional language. Our claim is therefore that the primes of phonological representation, their properties and number should only reflect categories of linguistic competence and not account for data belonging to the external/extensional dimension of language.

References

- Backley, Phillip. 1993. "Coronal: the undesirable element". *Working Papers in Linguistics* 5: 301-323.
- Backley, Phillip. 2011. *An Introduction to Element Theory*. Edinburgh: Edinburgh University Press.
- Backley, Phillip. 2012. "Variation in element theory". *Linguistic Variation* 12.1: 57-102.
- Backley, Phillips & Kuniya Nasukawa. 2009. "Headship as melodic strength". In Nasukawa & Backley (eds) *Strength relations in phonology*, pp. 47-77.
- Charette, Monik and Asli Göksel. 1996. "Licensing constraints and vowel harmony in Turkic languages". *SOAS Working Papers in Linguistics and Phonetics* 6: 1-25.
- Harris, John and Geoff Lindsey. 1995. "The elements of phonological representation". In Jacques Durand and Francis Katamba (eds.) *Frontiers of Phonology: Atoms, Structures and Derivations*. Harlow: Essex: Longman: 34-79.
- Jensen, Sean. 1994. "Is ? an Element? Towards a Non-segmental Phonology". *SOAS Working Papers in Linguistics & Phonetics* 4: 71-8.
- Kaye, Jonathan. 2001. Working with licensing constraints. In *Constraints and Preferences*, Katarzyna Dziubalska-Kolaczyk (ed.), 251-268. Berlin & New York: Mouton de Gruyter.
- Kaye, Jonathan. 2005. "GP, I'll have to put your flat feet on the ground". In H. Broekhuis *et al.* (eds.) *Organizing Grammar. Studies in Honor of Henk van Riemsdijk*. Berlin: Mouton de Gruyter: 283-288.
- Pöchtrager, Markus A. 2006. *The Structure of Length*. Dissertation zur Erlangung des Doktorgrades der Philosophie aus dem Fachgebiet Sprachwissenschaft, Universität Wien.
- Reiss, Charles. 2012. "Towards a bottom-up approach to phonological typology". In Anna Maria Di Sciullo (ed.), *Towards a Biolinguistic Understanding of Grammar: Essays on Interfaces*, Amsterdam: John Benjamins: 169-191.
- Scheer, Tobias. 1999. "A theory of consonantal interaction". *Folia linguistica* 32.3-4: 201-237.
- UPSID (*UCLA Phonological Segment Inventory Database*). Compiled by Ian Maddieson and Kristin Precoda.