Locality Domains and Phonological C-Command Over Strings

Thomas Graf, Stony Brook University

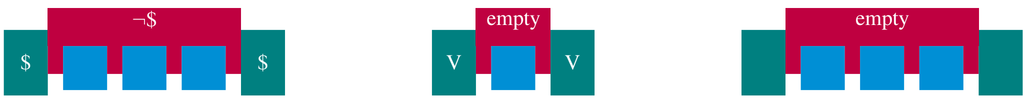
Keywords: computational linguistics, phonotactics, subregular complexity, long-distance dependencies

Overview A lot of recent work in computational phonology seeks to pinpoint the complexity of phonotactic dependencies from a formal perspective. Numerous mathematical classes have been proposed (see Heinz 2015), but Graf (2017) subsumes them all under the umbrella of *interval-based strictly piecewise* dependencies (IBSP). IBSP treats all phonological structures as strings that are subject to non-local constraints. Every local dependency is reanalyzed as a non-local constraint that is restricted to a locality domain of bounded size. If all phonotactic dependencies are IBSP, then we should only encounter phenomena that are either local or non-local, but not both. I show that this prediction is not borne out: several attested phenomena combine local and non-local information. However, a more relaxed notion of locality domains allows IBSP to produce these patterns while avoiding overgeneration. The make-up of these relaxed locality domains bears a striking resemblance to c-command in syntax, supporting the idea that phonology and syntax may be closely related on a computational level (Graf and Heinz 2015).

Non-local constraints Consider a locally unbounded phenomenon like sibilant harmony in Samala, where no word may contain sibilants that disagree in anteriority, irrespective of their relative distance. We can express this by the two inviolable markedness constraints *s···ʃ and *ʃ···s, or more succinctly as $^*[S,\alpha\,\text{ant}]\cdots[S,-\alpha\,\text{ant}]$ (where S is the feature specification of a sibilant). These constraints will rule out any string that contains both [s] and [ʃ], no matter how far the two are apart. Such non-local constraints are called *strictly piecewise* (SP). SP constraints only apply to a bounded number of segments, but they have no notion of locality, which makes them a bad fit for many phonotactic dependencies.

The locality problem of SP is illustrated by unbounded tone plateauing, where no low tone (L) may occur inside an interval spanned by two high (H) tones in the same word. So HLLLL, LLLLH, HHHHH, and LLLLL are well-formed, but HLLLH and LHLHH are not. One might expect that the ill-formed strings can be blocked with the SP-constraint *H···L···H. But this works only if the phonological representation consists of a single word. Since processes like tone sandhi often operate across the boundaries of phonological words, it seems more likely that the *H···L···H constraint would have to apply to a multi-word representation. But then it would also discard well-formed multi-word strings like $HLLLL**$**LLLLH$ — a case of overapplication that is due to ignoring the intermediate word boundary **$**.

IBSP As a remedy to the overapplication problem, Graf (2017) enriches SP with interval-like locality domains. Every locality domain is specified by a so-called k-val, which describes the **left and right edge** of the interval, the k **open slots** that are to be regulated, and the **fillers**. The word domain for unbounded tone plateauing, for instance, is represented by the 3-val in (1a). Any three segments x, y, and z then are



(a) Word domain *(b)* Intervocalic domain *(c)* 3-segment domain

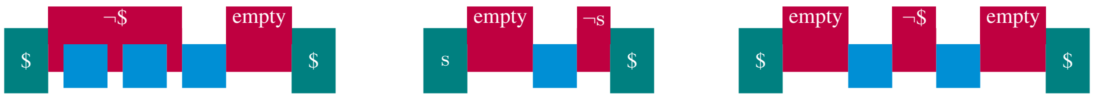
Figure 1: IBSP locality domains

subject to the SP-constraint *H···L···H only if it is possible to instantiate the 3-val in the string such that x, y, and z correspond to the open slots. For the word domain, this means that x, y, and z must occur between some word edges markers and may be separated by an arbitrary number of filler segments (including none at all) as long as said fillers are not word edges. Therefore x, y, and z can fill the open slots of the word domain only if they all belong to the same word, and consequently unbounded tone plateauing no longer applies across word boundaries. The non-local constraint has been locally restricted.

Power and limits of IBSP IBSP can also use k-vals to model locally bounded processes. Intervocalic voicing, for instance, is the combination of the 1-val (1b) with the constraint $^*[-\text{voice}]$. Alternatively, it can be handled by the constraint $^*\text{V}[-\text{voice}]\text{V}$ restricted to the domain (1c), which picks out any three adjacent segments. The strict locality of these domains arises from the ban against any fillers between the edges and the open slots. But this predicts that locality domains either enforce adjacency between all segments in a k-val, or none of them — phonotactic dependencies should never mix the two.

Local non-local dependencies At least four phenomena combine local with non-local information: non-final RHOL in Eastern Cheremis and Dongolese Nubian, local dissimilation in Samala, tone spreading in Copperbelt Bemba, and Yaka harmony domains. Only the first two are discussed here for space reasons.

Non-final RHOL (Hayes 1995) refers to a particular stress assignment pattern: primary stress falls on the rightmost heavy syllable in a non-final position, and otherwise on the leftmost light syllable. Hence ĹLLL, LLH́L, LLH́H, H́LLL, and ĹLLH are well-formed, but not ĹLHL or LLLH́. This pattern is not IBSP: Assume a string representation with only tone-bearing units H and L. To ensure that Ĺ only occurs in words without a viable heavy syllable, we have to look at the entire word, so fillers must be allowed between open slots. But in order for a heavy syllable to be a viable carrier of primary stress, it must not be word-final, which means that it must not be in an open slot that is immediately adjacent to the right word edge. Therefore the k-val would have to be of the form in (2a), which is not allowed for IBSP. With this locality domain, non-final RHOL is jointly enforced by the three constraints $^*\acute{X}\cdots H\cdots X$, $^*X\cdots\acute{L}\cdots X$, and $^*X\cdots X\cdots\acute{X}$ (X matches both L and H). The requirement for exactly one primary stress (i.e. culminativity) is already IBSP and thus easy to enforce independently.



(a) Non-Final RHOL domain *(b)* Samala dissimilation domain *(c)* First-last domain

Figure 2: Extended IBSP locality domains

Another example comes from Samala (McMullin 2016), where [sn], [sl] and [st] are forbidden unless there is an [s] somewhere to the right. Again this pattern mixes local and non-local information. The most compact account is provided by the locality domain (2b) in combination with the constraints *n, *l, and *t. This prevents those segments from being immediately preceded by [s] without also being followed by [s].

Overgeneration and c-command If IBSP is to be an empirically viable model of natural language phonotactics, it must allow for locality domains that combine local and non-local information. But this also introduces an undesirable amount of overgeneration. For instance, the locality domain (2c) can be combined with the constraint $^*[S,\alpha\,\text{ant}]\cdots[S,-\alpha\,\text{ant}]$ to produce a variant of Samala sibilant harmony that only holds between the first and the last segment in a word. This kind of first-last harmony is unattested.

The k-val for first-last harmony differs from the others in that it allows the adjacency requirements to be discontinuous. Intuitively, the domain it describes fits the pattern *local* · *non-local* · *local*, whereas the k-vals for attested phenomena are of the form $local^n \cdot non\text{-}local$ or $non\text{-}local \cdot local^n$ ($n = 1$ for (2a) and (2b) but is larger for Copperbelt Bemba and Yaka). Such patterns are also found in syntax: X c-commands Y iff Y is reflexively dominated by a sister of X; the sister-of relation is local, whereas reflexive dominance is a non-local relation, so c-command can be viewed as an instance of *local* · *non-local*. Other relations such as m-command and sub-command also fit this pattern. By contrast, there seem to be no instances of the discontinuous *local* · *non-local* · *local* pattern in syntax. Apparently the interaction of local and non-local information in both phonology and syntax is limited to c-command-like patterns, supporting Graf and Heinz's (2015) conjecture that phonology and syntax face comparable computational limitations.

Conclusion Graf (2017) models phonological dependencies via non-local constraints that are restricted by locality domains. In his IBSP approach, expressivity is crucially dependent on what shapes locality domains may assume. I have argued that these locality domains must be able to combine local and non-local information in a fashion resembling c-command in syntax. More powerful combinations induce undesirable overgeneration. Future research will have to explain, though, why discontinuous adjacency requirements should be disallowed for locality domains in phonology as well as syntax.

Graf, T. 2017. The power of locality domains in phonology. To appear in *Phonology*. ● **Graf**, T., and J. Heinz. 2015. Commonality in disparity: The computational view of syntax and phonology. Talk given at GLOW 2015. ● **Hayes**, B. 1995. *Metrical stress theory*. Chicago University Press. ● **Heinz**, J. 2015. *The computational nature of phonological generalizations*. Ms., UD. ● **McMullin**, K. 2016. *Tier-based locality in long-distance phonotactics: Learnability and typology*. PhD thesis, UBC.