Universal and Language Particular Constraints in OT-LFG

K P Mohanan and Tara Mohanan
National University of Singapore

Abstract: Optimality Theory claims that all typological variations can be explained in terms of language particular differences in the ranking of universal constraints. Based on patterns in verb morphology, and the phenomenon of OCP, this paper argues for another source of typological variation, namely, that of language particular manifestations of universal constraint cores, or constraint schemas, and thereby supplementing the mechanism of constraint ranking with that of constraint generation.

1. The nature of constraints in OT

The expression of the regularities in a system involves the statement of laws (rules, constraints, principles, conditions) and of their interaction. In the 'Aspects' theory of syntax, regularities of language structure were stated as language particular laws, and their interactions were expressed in terms of ordering. In Lexical Functional Grammar (Bresnan 1982, 2001), these regularities were stated as a combination of universal and language particular laws, and their interactions in terms of structure building. In Optimality Theory (Prince and Smolensky 1993), all regularities are stated as universal laws, and their interactions are expressed in terms of language particular ranking.

In adopting the position that all laws in human languages are universal, OT makes the further claim that all structural differences between languages can be deduced from the combination of universal constraints and language particular specifications of (a) the ranking of the constraints, and (b) the idiosyncratic properties of lexical items. The OT enterprise is to show that apparent complexity in human languages is the result of the interaction of a number of simple ingredients, namely, the universal constraints.

The purpose of this paper is to suggest that while this claim is clearly an attractive and desirable one as a methodological guiding principle, it may also be too strong to be taken as a theoretical claim. There exist structural differences within and across languages that nudge us to relax the OT claim, and allow for the expansion of the language particular possibilities. In place of a conception of inviolable and violable universal constraints, we argue for a conception of invariant and variable universal constraint cores to express regularities of language structure. Such a modification of the theory would involve universal constraints that are underspecified, such that the language particular component would 'fill in' the underspecified information, in addition to providing the ranking of the constraints and the idiosyncratic properties of lexical items.

Our goal is to argue for language particular constraints that are built out of the universal core, to illustrate how fully specified constraints can be generated from an underspecified core, and to show that in addition to constraint ranking, constraint generation is a source of typological variation.

2 Constraints in Malayalam verb morphology

Regularities in morphological structure constitute the most serious challenge to the claim that all constraints are universal. It is fairly obvious that there exist language particular aspects to morphological structure that are not necessarily a function of constraint ranking. Such differences, however, have generally been assumed to stem from the idiosyncratic properties of language particular morphemes. In what follows, we will examine this assumption on the basis of a case study of the verb morphology of Malayalam, and demonstrate that given certain assumptions of what counts as a universal constraint, the claim that all constraints are universal is untenable.

2.1 Word-internal c-structure and f-structure

We begin with an example that illustrates a multi-stem sequence in Malayalam verbs:

(**1**)   *kuTTi*   oo*TikkoNTirikkaNamaayirunnu*.
          child   should have continued to run
          The child should have kept running.

The morphological break up of the verb in (1) can be given as (2a), with the word internal c-structure as given in (2b), and the corresponding f-structure as in (2c): [1], [2]

(**2**)   a.   ***ooTikkoNTirikkaNamaayirunnu***   morphological break up of the verb
          ooT-i-koN-T-irikk-aNam-aa-i-irun-nu
          run-**PA**-*KOL*-**PA**-*IRIKK*-**MOD**-*AA*-**PA**-*IRIKK*-**PA**
          run - in the process of - should have been

b.

word internal c-structure[3]

---
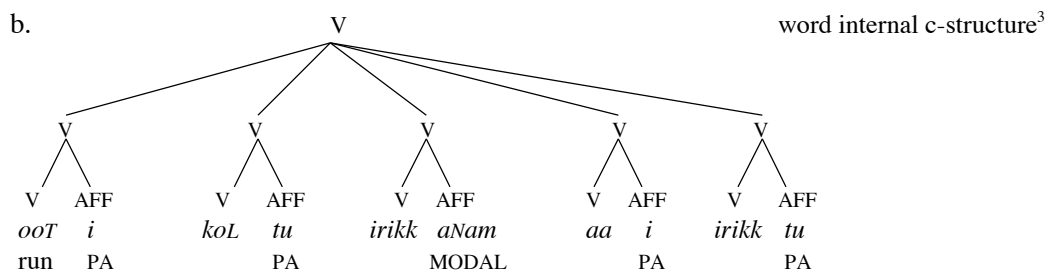
[1]   The following abbreviations are used in the glosses:

      PA: Past   PR: Present   FU: Future
      MOD: Modal   NF: Non-finite   REL: Relative clause marker

The forms glossed in upper case italics as *KOL, IRIKK,* and *AA*, are grammaticalized counterparts of the verb stems that mean 'fit', 'sit', and 'be/become' respectively.

[2]   See Asher & Kumari (1997) for a description of the structure of Malayalam, and Mohanan & Mohanan (forthcoming) for an extensive discussion of the facts and analysis of Malayalam verb morphology.

[3]   What we have called word-internal c-structure in (2b) (also Mohanan and Mohanan forthcoming) is the same as what the recent LFG literature refers to as m-structure (Butt et al 1996, Sadler & Spencer 2001, Frank & Zaenen 2002).

c.

$$\begin{bmatrix} \text{PRED 'run' (SUBJ)} \\ \text{TENSE: PAST} \\ \text{MODAL: OBLIGATION} \\ \text{PROG: +} \\ \text{CONT: +} \end{bmatrix}$$

word internal f-structure

Representations of the type illustrated in (2) raise the following questions. What are the regularities governing word-internal c-structure in Malayalam? How do we predict these regularities from a set of statements in UG? What are the regularities governing the correspondence between word-internal c-structure and f-structure? For instance, the verb form in (1) contains four occurrences of morphological past ((2b)) corresponding to a single occurrence of syntactic past ((2c)). Should the regularities in such correspondences be expressed as universal constraints, as language particular constraints, as a combination of the two, or as stipulations on lexical items? In classical LFG, the pairing between c-structure and f-structure in the phrasal domain is handled by annotations on language particular phrase structure rules. Given that in OT-LFG, language particular rules are replaced by universal constraints, the device of annotations needs re-thinking.

2.2 Tense and negation

Malayalam has a three-way tense system, with past, present, and future tenses, as illustrated in (3). The nonfinite affixes in the language are illustrated in (4):

(**3**) a. *ooTi*   b. *ooTunnu*   c. *ooTum*
              run-PA   run-PR   run-FU
              'ran'   'runs'   'will run'

(**4**) a. *ooTaan*   b. *ooTuka*   c. *ooTaaR-*
              run-NF1   run- NF2   run- NF3
              'to run'   'running'   'about to run'

The tense affixes interact with the negation marker *illa*, as shown in (5):

(**5**) a. *ooTiyilla*   b. *ooTunnilla*   c. * ***ooTumilla***   d. *ooTilla*
              run-PA-NEG   run-PR-NEG   **run-FU-NEG**   run-NEG
              'didn't run'   'isn't running'   'won't run'

Notice that (5c), where the negation marker follows the future tense affix, is unacceptable; future negation is expressed by (5d), without a future tense marker. It should be noted that a verb form without a tense marker (e.g., *ooT*) cannot be the sole verb in a sentence, except in the context of negation and that the V+NEG form without the future afix cannot be interpreted as anything other than future tense. These observations about the interaction between tense and negation can be summarized as the generalizations in (6):

(**6**) a. V+FUT+NEG is an illformed c-str representation. (c-structure gap: (5c))

  b. V+NEG in c-str corresponds to [PRED, FUT] in f-str. (c-str/f-str mismatch: (5d))

Thus, the negative marker *illa* imposes c-structure restrictions on the verb, not reflected in its f-structure ((6a)), which calls for statements of mismatch between c-structure and f-structure ((6b)).

The generalizations in (6) can be expressed in terms of the lexical specifications on the negative marker, as in (7), together with the constraints in (8) and (9):

(**7**) *-illa* 'NEG': (a) mother = [+tense]

  (b) sister = [–tense]

(**8**) a. Every c-structure feature complex must be associated with a corresponding f-structure feature complex, and vice versa.

  b. An unassociated f-structure carrying a paradigmatic contrast is paired with an available well-formed unassociated c-structure.

(**9**) a. All elements of a paradigm must be expressed.

  b. Paradigmatic contrasts must be preserved.

We assume that (7a) and (7b) are language particular stipulations on lexical entries, while (8)-(9) are universal statements. (7b) correctly disallows the form in (5c). (7a) provides the interpretation of V+NEG as tensed.[4] Given the independently available past and present forms in (5a, b), the constraints in (9) derive the interpretation of (5d) as future negative.

2.3 Tense and relative clauses

Like negation, the relative clause marker *–a* in Malayalam imposes morphological restrictions on tense marking. The relative clause construction is illustrated in (10b, c), embedding (10a):

(**10**) a. *kuTTi*   *weegam*   *ooTi / ooTunnu*
  child   fast   run-PA / run-PR
  the child who ran fast

  b. *weegam*   *ooTiya*   *kuTTi*
  fast   run-PA-**REL** child
  the child who ran fast

  c. *weegam*   *ooTunna*   *kuTTi*
  fast   run-PR-**REL** child
  the child who runs/will run fast

_____

[4]   The specification "mother = [+tense]" in (7a) can translated as ↑tense= +. LFG doesn't currently have a notation for 'sister = X'.

Observe that unlike (10b), (10c) is ambiguous between present and future tense readings. Related to this is the fact that the relative clause suffix *-a*, like the negative suffix *-illa*, cannot co-occur with the future tense suffix *-um*, as shown by (11a). Nor can it co-occur with a tenseless verb, as shown by (11b). In this, it is unlike the negation marker:

(**11**) a. * *ooTuma* b. * *ooTa*

run-FU-**REL** run-**REL**

Thus, the two constructions differ in how the c-structure gap is filled. While the future negative is expressed by a tenseless verb combining with the NEG marker, the future relative clause has no distinct form: the present tense verb form expresses both present and future ((10c)). These generalizations can be captured in terms of the lexical stipulations in (12) and supplementing the universal constraints in (8) and (9) with the default constraint in (13):

(**12**) *-a* 'COMP': (a) mother = [+tense]
  (b) sister = [+tense, –fut]

(**13**) Default: c-structure features are identical to the f-structure features. (faithfulness)

2.4 Multiple morphological tenses

In Malayalam, aspectual, modal, and voice contrasts are expressed by a sequence of verb stems within a single verb form. This results in more than one tense affix appearing in a verb, as illustrated in (14):

(**14**) a. *ooTumaayirunnu* * *ooTiyaayirunnu* * *ooTunnaayirunnu*
  ooT-um-aa-i-ir-ikk-tu ooT-i-aa-i-ir-ikk-tu ooT-unnu-aa-i-ir-ikk-tu
  run-**FUT**-*AA*-**PA**-*IRIKK*-**PA** run-**PA**-*AA*-**PA**-*IRIKK*-**PA** run-**PRES**-*AA*-**PA**-*IRIKK*-**PA**
  used to run

  b. *ooTukayaayirikkum* *ooTaaraayirikkum*
  ooT-uka-aa-i-ir-ikk-um ooT-aaR-aa-i-ir-ikk-um
  run-**INF**-*AA*-**PA**-*IRIKK*-**FUT** run-**NF**-*AA*-**PA**-*IRIKK*-**FUT**
  will be in the process of running will be about to run.

The verb form in (14a) carries the future tense affix, and two instances of the past tense affix. The unacceptable forms in (14a) show that the non-final future marker cannot be replaced by any of the other tense markers. Notice also that the non-final future affix in c-structure does not contribute syntactic FUT to the f-structure. Likewise,(14b) illustrates that the non-final past tense marker does not contribute syntactic PAST to the f-structure. That only the final tense marker contributes the syntactic tense is true of all verb forms in Malayalam.

The examples in (15)-(18) illustrate the use of various affixed verb stems to express a range of durative meanings.

(**15**) a. *kuTTi*   *ooTunnu.*
child-N   run-**PRES**
The child runs.

  b. *kuTTi*   *ooTi.*
child-N   run-**PA**
The child ran.

(**16**) a. *kuTTi*   *ooTukayaakunnu.*
child-N   run-INF-*AA*-**PRES**
The child is running.

  b. *kuTTi*   *ooTukayaayirunnu.*   (cf: *ooTukayaayi.*
child-N   run-**INF**-*AA*-**PA**-*IR*-**PA**   run-**INF**-*AA*-**PA**
The child was running.   is about to run.)

(**17**) a. *kuTTi*   *ooTikkoNTirikkunnu.*
child-N   run-**PA**-*KOL*-**PA**-*IR*-**PRES**
The child keeps running.

  b. *kuTTi*   *ooTikkoNTirunnu.*
child-N   run-**PA**-*KOL*-**PA**-*IR*-**PA**
The child kept running.

(**18**) a. *kuTTi*   *ooTikkoNTirikkukayaakunnu.*
child-N   run-**PA**-*KOL*-**PA**-*IR*-**INF**-*AA*-**PRES**
The child is "keeping on running".

  b. *kuTTi*   *ooTikkoNTirikkukaayaayirunnu.*   (cf. *ooTikkoNTirikkukaayaayi.*
child-N   run-**PA**-*KOL*-**PA**-*IR*-**INF**-*AA*-**PA**-*IR*-**PA**   run-**PA**-*KOL*-**PA**-*IR*-**INF**-*AA*-**PA**)
The child was "keeping on running".

The examples in (15)-(18) further show mismatches in c-structure/f-structure correspondences, which warrant a closer look. Note that as in the case of negation and relative clauses, the c-structure features of these verb forms are not automatically projected into f-structure. To unearth the constraints governing the correspondences, we must juxtapose the two sets of features in (15)-(18) as in (19)-(22) respectively, where the meanings of past (time), pres(ent time), prog(ression), and cont(inuation) are associated with f-structure contrasts. The labels *AA*, *KOL* and *IR* (the roots of the grammaticalized verbs 'be', 'fit', and 'sit' respectively) below are place holders for appropriate c-structure features:

		INPUT		OUTPUT	

(19) a. pred pres → pred pres   f-str

                                               V-PRES   c-str

       b. pred past → pred past   f-str

                                                 V-PAST   c-str

(20) a. pred prog pres → pred prog pres   f-str

                                               V-INF-*AA*-PRES   c-str

       b. pred prog past → pred prog past   f-str

                                               V-INF- *AA* -PAST-*IR*-PAST   c-str

(21) a. pred cont pres → pred cont pres   f-str

                                               V-PAST-*KOL*-PAST- *IR* −PRES   c-str

       b. pred cont past → pred cont past   f-str

                                               V-PAST-*KOL*-PAST- *IR* −PAST   c-str

(22) a. pred cont prog pres → pred cont prog pres   f-str

                                               V-PAST-*KOL*-PAST- *IR* -INF- *AA* −PRES   c-str

       b. pred cont prog past → pred cont prog past   f-str

                                               V-PAST-*KOL*-PAST- *IR* -INF- *AA* -PAST- *IR* −PAST   c-str

The roots *KOL* and *AA* IN (19)-(22) carry the meanings of continuation and progression respectively. The root *IR*, an expletive in these forms, serves a purely morphological function. The morphological break up of the output candidates of these examples, given in (23)-(26) below, separates the meaning-carrying elements (bold face) from the expletive morphology:

(23) a. [**V**-**PRES**]$_{pres}$
       b. [**V**-**PAST**]$_{past}$

(24) a. [**V**-INF] [*AA*-**PRES**]$_{prog\text{-}pres}$
       b. [**V**-INF] [*AA*-PAST-*IR* -**PAST**] $_{prog\text{-}past}$

(25) a. [**V**-PAST [*KOL*-PAST-*IR*- **PRES**]$_{cont\text{-}pres}$
       b. [**V**-PAST] [*KOL*-PAST- *IR*- **PAST**]$_{cont\text{-}past}$

(26) a. [**V**-PAST] [*KOL*-PAST-*IR*-INF]$_{cont}$ [*AA*-**PRES**]$_{prog\text{-}pres}$
       b. [**V**-PAST] [*KOL*-PAST-*IR*-INF]$_{cont}$ [*AA* -PAST-*IR*-**PAST**]$_{prog\text{-}past}$

We informally state in (27) the restrictions that govern the choice and sequencing of elements in the verb morphology of Malayalam as exhibited in (15)-(18):

(**27**) a.   A verb stem is in the [PAST] tense form before another verb stem.

     b.   A verb stem is in the [INF] form before a progressive verb stem.

     c.   Within the continuative, the *KOL* stem requires an *IR* stem after it.

     d.   Within the progressive past, the *AA* stem requires an *IR* stem after it.

The restrictions in (27a, b) are not lexical specifications on individual morphemes, but general conditions on how stems with certain c-structure features combine. In other words, they are constraints abstracted away from particular morphemes. In contrast, (27c) and (27d) may be viewed as distributional restrictions of specific morphs (strict subcategorization), analogous to restrictions such as "–*ity* requires an adjective as its left sister" and "*un*— requires an adjective or verb as its right sister" in English morphology. Alternatively, they may be viewed as constraints on universally specified lexical classes that *KOL. AA* and *IR* belong to (*a la* Bresnan and Nikitina 2003).

The regularities expressed in (27) are not unlike the familiar constraints on auxiliaries that we find in English such as those in (28):

(**28**) A verb stem is in:

     a.   the bare infinitival form after a modal verb.

                               e.g., *Sue will go/*goes/*went/*gone/*going.*

     b.   the –*ing* form after a progressive verb stem.

                               e.g., *Sue is going/*go/*goes/*went/*gone.*

     c.   the –*en* form after a perfective or passive form.

                               e.g., *Sue has gone/*go/*goes/*went/*going.*

To the extent that the specific restrictions in (27) are found only in Malayalam, and those in (28) are found only in English, neither of them can be legitimately regarded as universal constraints.

Examples of word-internal c-structure constraints that are f-structurally unmotivated are found not only in the verbal system but also in the nominal system of Malayalam. The examples in (29) below illustrate the phenomenon in the case system, where the form of the nominal that case marking attaches to is determined by the final phonological segment of the stem:

(**29**)

		NOM	ACC	DAT	COM	GEN	INSTR
a.	'rat'	eli	eliye	elik'k'∂	eliyooT∂	eliyuTe	eliyaal
b.	'flower'	puu	puuwine	puuwin∂	puuwinooT∂	puuwinte	puwinaal
c.	'tree'	maram	marattine	marattin∂	marattinooT∂	marattinte	marattinaal

In (29a) is given a straightforward example where the case marking attaches to the bare nominal stem, which is identical to the nominative form. The special morphological constraints needed for the paradigms in (29b, c) can be stated as (30a, b) respectively:

(**30**) a.   When the stem ends in *u*, the morph -*in* must be attached to the host of case marking.

     b.   When the stem ends in *m*, the morphs -*tt+in* must be attached to the host of case marking.

The constraints in (30) apply to the class of (non-nominative) case affixes.

2.5 Word-internal c-structure: language particular or universal?

In the context of OT, it would be useful to distinguish between constraints that are ACTIVE and those that are SUPPRESSED. For any given input to which a particular constraint is applicable, the constraint is active when there is no outranking constraint that conflicts with it, and is suppressed if a competing constraint outranks it. Given this distinction, a constraint is GLOBALLY ACTIVE in a language if, for all inputs, it has no outranking constraint that conflicts with it. It is GLOBALLY SUPPRESSED if, for all inputs, a competing constraint outranks it.

In the previous section, we took the position that the c-structure constraints in (27) and (28) are not universal, on the grounds that each of them is obeyed only in that particular language. The same remarks would apply also to (30). Technically speaking, it is indeed feasible to postulate them as universal constraints that are outranked by either a faithfulness constraint or a complementary constraint in every other language, such that they are globally suppressed in every language but one. Given the freedom to postulate such constraints, the claim that all constraints are universal becomes empirically vacuous. To avoid such vacuity, we suggest a methodological guideline on what can be legitimately considered a universal:

(31)   To qualify as a universal, a constraint should be active in at least a few unrelated languages.

The effect of (31) is that to qualify as a universal constraint, the pattern it expresses should be cross-linguistically recurrent. The constraints in (27), (28), and (30) are not universal unless they satisfy (31).

3 Towards a theory of constraints in OT

OT is essentially a theory of constraint interaction. As the discussion in section 2 suggests, a theory of constraint interaction needs to be supplemented by a theory of constraints that tells us what kinds of constraints are legitimate. In this section, we argue that such a theory should also include a sub-theory of constraint generation, and sketch the rudiments of a theory that generates language particular constraints as variable manifestations of a universal constraint core.

3.1 The concept of universals

The methodological requirement in (30) calls for a clarification of the concept of universals as we understand it in current linguistic theory. First, we need to recognize different degrees of 'universality' in the observed cross-linguistic regularities. Some universals are cross-linguistically invariant, while others exhibit variability in their instantiation. Among the invariant patterns, we find exceptionless *absolute universals* (e.g., disjoint reference), *strong universals* that hold in nearly all languages, though with a few exceptions (e.g., subject condition, violated, for instance, in Hindi), and *recurrent universals* that repeatedly found across languages (e.g., clause-bound anaphora).

Universals with variable instantiations are of two types. In some, variation is a matter of choice from a small set of options provided in the principle itself (the leading idea in the Principles and Parameters program). An example of *principle-and-parameters universals* is the constraint that the antecedent of a reflexive must be the most prominent element along dimension x. The relevant dimension for 'prominence' may be argument structure (logical subject, as in Hindi and Marathi), grammatical function structure (grammatical subject, as in Hindi, Malayalam, Japanese, Malay, and so on), discourse syntax (topic, as in Malay), or discourse semantics (logophoric center, as in Malayalam and Japanese).

In other variable instantiations, the variability lies in the manifestation of a single abstract archetype, where a single constraint schema is manifested as a number of constraints with family resemblances. An example of *archetype-and-manifestations universals* is the passive construction, which exhibits a universal core (the logical subject is not the grammatical subject), but has variable manifestations. For instance, in German and English, the logical subject is demoted, but not in Tagalog. German allows passives without a grammatical subject (impersonal passives), while English and Tagalog do not. Passives in Japanese and Mandarin have an adversity specification, but not those in English or Malayalam.

The typology of universal constraints sketched above suggests that there may be two sources for typological variation across languages. Some variations stem from the interaction of constraints, while others stem from the instantiation of a universal. OT seeks to use a single mechanism, that of constraint ranking, to derive all typological variation. In the light of the above discussion, the OT position calls for closer scrutiny.

3.2 Deriving language particular manifestations

If we accept the idea of universal archetypes and language particular manifestations of constraints, it follows that while a constraint is language particular in the sense that it is observed only in one language, or a set of related languages, it may nevertheless be derived from a universal schema. Let us re-examine the facts of verb morphology in Malayalam from this perspective.

Let us first take the association between c-structure and f-structure. Given the constraints internal to the c-structure of Malayalam verbs, the f-structure interpretation of the relevant syntactic features can be made to follow from the universal principles of c-structure/f-structure correspondence given in (8), (9) and (13). What we need, in addition, are the universal principles that identify the location of the syntactic finite tense, which we state as (32a, b):

(**32**) a.　　The verb sequence in a finite clause has one and only one syntactic FINITE TENSE.

　　　 b.　　Syntactic FINITE TENSE is borne by the unit at the edge.　　　[edge: left/right]

The consequence of (32a, b) is that verbs that are not at the edge are NON-FINITE. Typological variation deriving from (32), and the faithfulness constraint in (13), can be illustrated as follows:

	Malayalam	English
(32a)	not violated	not violated
(32b)	right edge	left edge
(13)	violated	not violated

Is it possible to pursue a similar approach to the constraints internal to c-structure? As remarked earlier, it is possible to take the position that regularities governing the distribution of individual morphs are expressed as stipulations on morphs, not as constraints. If so, the regularities in (27c, d) can be expressed as language particular stipulations on morphs (as in the case of the stipulation that –*ity* in English requires an adjective as its left sister).

However, (27a, b), and (28b, c), and (30a, b) cannot be dealt with as lexical stipulations; they must be expressed as constraints, or as patterns derived from the interaction of constraints. In the absence of a demonstration that these patterns are expressible as universal constraints (subject to (31)), or are derivable from the interaction of universal constraints, they are a counterexample to the claim that all constraints are universal. An alternative would be to derive the patterns from a universal constraint core; we have yet to determine whether this is feasible.

3.3 Underspecified constraints and constraint generation

Let us take a closer look at the strategy of deriving language particular constraints from a universal schema. Consider the following ranking of constraints proposed for the English dative alternation in Bresnan and Nikitina (2003):

(**33**) OO-PRIMACY>> FAITH-LATINATE (REC) >>FAITH$_{\{yell\}}$, FAITH$_{\{drag\}}$ >> HARMONY (1,2)

>> FAITH$_{\{fax\}}$, FAITH$_{\{throw\}}$ >> … >> FAITH $_{\{give\}}$ (REC) >> *STRUCT

Of the constraints referred to in (33), FAITH-LATINATE (REC), FAITH$_{\{yell\}}$, FAITH$_{\{drag\}}$, FAITH$_{\{fax\}}$, FAITH$_{\{throw\}}$ and FAITH $_{\{give\}}$ (REC) are variants of the same constraint that differ only in the domain of application of the constraint, specified as a lexical class. Bresnan and Nikitina express the constraint schema that underlies the different constraints in (33) as (34):

(**34**) FAITH (REC): Express the recipient role of a verb with distinct marking (case or adposition).

The redundancies in the formulation of the constraints in (33) can be eliminated by factoring out the domain specification of the constraint from the underspecified core constraint in (34):

(**35**) Universal constraint schema   Language particular constraints
        (underspecified)   (fully specified)
         FAITH (REC)   →   FAITH-LATINATE, FAITH$_{\{yell\}}$, FAITH$_{\{drag\}}$, and FAITH $_{\{give\}}$ …

We show in Mohanan & Mohanan (2003) that the unity underlying the diverse manifestations of archetypal phonological patterns like place assimilation, voicing assimilation, and intervocalic lenition are best captured by separating their universal core from the variable language particular manifestations. We also propose a mechanism for constraint generation that adds the specifications of locus (undergoer), trigger, domain, and outcome value to the constraint schema, thereby deriving the fully specified language particular constraints from the universal core.

Place assimilation of nasal stops, for instance, is found in almost all human languages. In languages like English, all stops assimilate, including oral stops; in languages like Malayalam, but only nasals (not oral stops) assimilate. In English, only coronals undergo assimilation in the domain of the foot, though both coronals and non-coronals undergo assimilation within a syllable; in Malayalam, both coronals and non-coronals undergo assimilation within a phonological phrase. The trigger in English must be non-coronal, whether in the domain of a foot or a syllable; in Malayalam, both coronals and non-coronals can be triggers. The unity and variability of place assimilation in these languages are expressed by the following analysis:

(**36**) <u>Universal core</u>:
Stop consonants (nasals and plosives) agree in their place of articulation with the following consonantal segment.

(**37**) <u>Language particular specifications</u>

	<u>locus</u>	<u>trigger</u>	<u>domain</u>	<u>language</u>
a.	[–nasal]	[+stop]	phon. phrase	Malayalam
b.	—	[–cor]	syllable	English
c.	[+cor]	[–cor]	foot	English

Having illustrated the idea of constraint generation with an example from phonology, we now turn to a similar example that spans phonology, morphology, and syntax.

3.4 OCP as a universal constraint schema

The Obligatory Contour Principle (OCP), originally proposed as a prohibition against adjacent identical tones (Leben 1973, Goldsmith 1976), and subsequently extended to other phonological elements (McCarthy 1986), has found further extensions into a prohibition against adjacent identical morphological and syntactic elements as well (Yip 1987, 1998, T. Mohanan1994, Fong and Anttila 2000), with the result that the core pattern of OCP can be stated as in (38):

(**38**) * Adjacent $\alpha\,\alpha$ in D.
(= Adjacent identical elements of the specified type are prohibited in the specified domain.)

An example of OCP in English that must make reference to morphological information is the prohibition against the co-occurrence of the plural and possessive *-(e)s*, illustrated in (39):

(**39**) a. *the child*   d. *the child's*
  b. *the children*   e. *the children's*
  c. *the boy's*   f. * *the boys's*

The ungrammaticality of (39f) would follow if we assume that OCP applies to the morph { z } in English:

(**40**) Locus of (38) in English: morph: *-z*

The disjunctive coordinator *–oo* 'or' and the conjunctive coordinator *–um* 'and' in Malayalam, both of which attach to the last word of every constituent they coordinate, exhibit a similar phenomenon. These coordinators are illustrated in (41b) and (41c) respectively:

(**41**) a. *kuTTi   ooTi*
        child   ran
        The child ran.

  b. *kuTTiyoo   ammayoo   ooTi*
      child-or   mother-or   ran
      The child or the mother ran.

  c. *kuTTiyum   ammayum   ooTi*
      child-and   mother-and   ran
      The child and the mother ran.

Given in (42) are examples of clefts in Malayalam, where (42b) involves a disjunction:

(**42**) a. kuTTiyaaN∂   ooTiyaat̲∂.
        child-is   run-it
        It was the child who ran.

  b. kuTTiyoo ammayoo aaN∂ ooTiyaat̲∂.
      child-or   mother-or is   run-it
      It was either the child or the mother who ran.

The morph *–oo* in Malayalam also functions as a *yes-no* interrogative marker when attached to the verb, as in (43a, b), where (43b) is a clefted question:

(**43**) a. *kuTTi   ooTiyoo?*
        child   ran-Q
        Did the child run?

  b. kuTTiyaaNoo   ooTiyaat̲∂?   c. * kuTTiyooaaN∂   ooTiyaat̲∂?
      child-is-Q   run-it   child-Q-is   run-it
      Was it the child who ran?

Consider now a clefted disjunctive interrogative construction, illustrated in (44a):

(**44**) a. kuTTiyaaNoo   ammayaNoo   ooTiyat∂?
          child-is-Q/or   mother-is-Q/or ran it
          Was it the child or the mother who ran?

      b. * kuTTiyaaNoo-oo   ammayaNoo-oo   ooTiyat∂
          child-is-Q-or   mother-is-Q-or ran it

The unacceptability of (44b) is evidence for the prohibition of the co-occurrence of the question marker –*oo* and the disjunction marker –*oo*.

The morph –*um*, a conjunctive coordinator illustrated in (41c), also functions as a universal operator in Malayalam, illustrated in (45b), where it signals the meaning 'all of X'. In the presence of an over quantifier 'all', as in (45c), the universal operator is obligatory, as shown by the ungrammaticality of (45d):

(**45**) a. naal∂   kuTTikaL
          four   children

      b. naal∂   kuTTikaLum
          four   children-∀
          All the four children.

      c. ellaa   kuTTikaLum
          all   children-∀
          All (the) children.

      d. * ellaa   kuTTikaL
          all   children

Now consider the interaction between the conjunction –*um* and the universal operator –*um*:

(**46**) a. ellaa kuTTikaLum   ellaa ammamaarum
          all   children- and/∀ all   mothers-and/∀
          All (the) children and all (the) mothers.

      b. * ellaa kuTTikaLumum ellaa ammamaarum
          all   children- and-∀ all   mothers-and-∀

Once again, the unacceptability of (46b) can be explained by the prohibition against the co-occurrence of the conjunction –*um* and the universal operator –*um*. The two prohibitions, illustrated in (44b) and (46b), follow from the language particular specifications in (47):

(**47**) Locus of (38) in Malayalam: (i)   morph: -*oo*
                                        (ii) morph: -*um*

The constraints in (40) and (47) hold on specific morphs, which are language particular units. Hence, the constraints themselves could not be universal. We must therefore assume that only the schema in (38), from which these constraints are generated, is universal.

A more complex instance of OCP is that Case OCP, recurrently found across languages. The generalization governing Case OCP in Hindi (T. Mohanan 1994) is as stated in (48):

(**48**)　　Hindi:　Identical case formatives in adjacent phonological words that are associated with participants of the same predicate are prohibited within a phonological phrase.

This generalization, which applies to all case morphs, can be expressed as in (49):

(**49**)　　Locus of (38) in Hindi:　case formative
　　　　　　　　　　　　　　　　　　　|
　　　　　　　　　　　　　　　　　argument
　　　　　　　　　　　　　　　　　　|
　　　　　　　　　　　　　　phonological word

　　　　　　Domain of (38):　　　f-structure clause
　　　　　　　　　　　　　　　　　　　|
　　　　　　　　　　　　　phonological phrase

In Japanese, unlike in Hindi, Case OCP applies only to the accusative case marker *-o* internal to an S node. Malayalam, unlike both Japanese and Hindi, exhibits no Case OCP. If we accept the methodological principle in (31), then Case OCP in Hindi, whether stated as an underived complex constraint in (48), or as a constraint derived from the combination of the simple underspecified core in (38) and the language particular specifications in (49), is not a universal constraint.

4. Concluding remarks

If we accept the analysis of verbal morphology and of OCP, we must acknowledge a source of typological variation other than the one due to differences in the language particular *ranking* of universal constraints, namely, variation due to differences in the language particular *specification* of the underspecified elements in universal constraint schemas. Adopting this position leads to the possibility of supplementing the device of *constraint ranking* with that of *constraint generation* in a theory that provides an account of the emergence of language particular constraints, thereby yielding typological differences between languages in terms of differences in the way constraints are assembled from universal ingredients. Such a theory would allow for two types of constraints, namely, (i) universal constraints, and (ii) language particular constraints that emerge from (are built out of) universal ingredients.

References

Asher, R. E. and T. C. Kumari. 1997. *Malayalam*. London and New York: Routledge.

Bresnan, Joan (ed). 1982. *The Mental Representation of Grammatical Relations*. Cambridge: The MIT Press.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Blackwell Publishers.

Bresnan, Joan and Tatiana Nikitina. 2003. On the gradience of the dative alternation. Ms. Stanford University.

Butt, Miriam, M. E. Niño, and F. Segond. 1996. Multilingual processing of auxiliaries in LFG. In D. Gibbon (ed.) *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference*. Bielefeld. 111-122.

Fong, Vivienne and Arto Anttila. 2000. The partitive constraint in Optimality Theory. *Journal of Semantics* 17.4:281-314. [Available on Rutgers Optimality Archive, ROA-416-09100]

Frank, Annette and Annie Zaenen. 2002. Tense in LFG: syntax and morphology. In Hans Camp and Uwe Reyle (ed.) *How we say WHEN it happens: Contributions to the theory of temporal reference in natural language*. Niemeyer, Tubingen.

Goldsmith, John. 1976. *Autosegmental Phonology*. Doctoral dissertation, MIT. Published 1979, Garland Press, New York.

Leben, Will. 1973. *Suprasegmental Phonology*. Indiana University Linguistics Club.

McCarthy, John. 1986. OCP Effects: Gemination and antigemination. *Linguistic Inquiry* 17:207-263.

Mohanan, Tara and K P Mohanan. 2003. Towards a Theory of Constraints in OT: Emergence of the not-so-unmarked in Malayalee English. [Available on Rutgers Optimality Archive, ROA-416-09100]

Mohanan, Tara and K P Mohanan. To appear. Multiple Tenses in the Malayalam Verb. In Sharon Inkelas and Kristin Hanson (eds) *The Nature of the Word: Essays in Honor of Paul Kiparsky*. Cambridge, Mass.: The MIT Press.

Mohanan, Tara.1994. Case OCP: A Constraint on Word Order in Hindi. In Miriam Butt, Tracy King, and Gillian Ramchand (eds). *Theoretical Perspectives on Word Order in South Asian Languages*. pp. 185-216.Stanford. California: CSLI Publications.

Prince, Alan and Paul Smolensky. 1993. Optimality Theory: constraint interaction in generative grammar. RuCCS Technical Report #2. Piscateway, NJ: Rutgers University Center for Cognitive Science.

Sadler, Louisa and Andrew Spencer. 2001. Syntax as an exponent of morphological features. In Geert Booij (ed.) *Yearbook of Morphology 2000*. Kluwer Academic Publishers. pp 71-96

Yip, Moira. 1988. The Obligatory Contour Principle and phonological rules: A loss of identity. *Linguistic Inquiry* 19:65-100.

Yip, Moira. 1998. Identity avoidance in phonology and morphology. In Steven G Lapointe, Diane K. Brentari, and Patrick M. Farrell (eds) *Morphology and its Relation to Phonology and Syntax*. Stanford: CSLI Publications. 216-246.