Pp 19–35 in T. Mark Ellison & James M. Scobbie (eds) 1993 **Computational Phonology**,
Edinburgh Working Papers in Cognitive Science **8**.
Available by anonymous FTP from scott.cogsci.ed.ac.uk.

An HPSG Approach to Sierra Miwok Verb Stems*

Ewan Klein

1 Introduction

Typed feature structures (Carpenter 1992) impose a type discipline on unification-based grammar formalisms. A partial ordering over the types gives rise to an inheritance hierarchy of constraints. As Emele and Zajac (1990) point out, this object-oriented approach brings a number of advantages to grammar writing, such as a high level of abstraction, inferential capacity, and modularity.

On the face of it, such benefits should extend beyond syntax, to phonology, for example. Although there have been some valuable efforts to exploit inheritance and type hierarchies within phonology (e.g. Bird 1990; Gibbon 1987), the potentials of typed feature terms for this area have barely been scratched so far. This paper takes a step in exploring the terrain by couching an analysis of templatic morphophonology within the framework of HPSG (Pollard and Sag 1987).

The data examined—Goldsmith 1990's presentation of Sierra Miwok verb stems—is deliberately minimal, since my main concern has been to determine the feasibility of the approach, rather than to carry out a detailed analysis.

2 Framework

The basic grammar object in HPSG is the feature structure of type *sign*. Objects of this type are constrained to have the following features defined:[1]

(1) $$\left[\begin{array}{ll}\text{PHON}: & \textit{phon}\\ \text{SYNSEM}: & \textit{synsem}\end{array}\right]_{\textit{sign}}$$

That is, *sign*s must contain the attributes PHON (phonology) and SYNSEM (syntax-and-semantics), and these attributes must take values of a specific type (*phon* and *synsem* respectively).

*I am grateful to Steven Bird, Tomaž Erjavec, and Jim Scobbie for their constructive attempts to make sense of this paper, and to Martin Emele, Martin Kay, Susanne Riehemann and Ivan Sag for stimulating discussions of the general issues involved.

[1]The feature constraints illustrated here and in the rest of the paper omit much of the detail developed in standard versions of HPSG. Note also that attribute value matrices like (1) are best regarded as constraints on feature structures, rather than feature structures themselves; see Pollard and Sag 1993 for more discussion of this issue.

The set of types is partially ordered. If σ is a subtype of τ (sometimes written $\sigma \leq \tau$), then σ obeys all the constraints that τ does. For example, *stem* is a subtype of *sign*; not only does *stem* inherit all the constraints obeyed by *sign*, but as we will see later, it adds some constraints of its own. Types which do not have any subtypes are called *minimal*.

A type declaration for *sign* defines it as the following disjunction of subtypes:

(2) $\quad$ *sign* = *root* ∨ *stem* ∨ *word* ∨ *phrase*

We can also assign constraints to the types *synsem* and *phon* which occurred as values in (1), as shown in the next examples:

(3)
$$ {}_{synsem}\begin{bmatrix} \text{CAT}: & cat \\ \text{SUBCAT}: & list \\ \text{SEM}: & semantics \end{bmatrix} $$

(4)
$$ {}_{phon}\begin{bmatrix} \text{CON}: & list \\ \text{VOW}: & list \\ \text{SKEL}: & list \end{bmatrix} $$

Later, we will see more examples of how types are declared and constrained.

A standard constraint on type hierarchies is that subsumption corresponds semantically to set inclusion; that is, the set of objects belonging to a type σ is a subset of the objects belong to all its supertypes. However, we can impose other conditions of varying degrees of strength on the interpretation of typed feature terms. Following (Manandhar 1993; Zajac 1992), I will adopt a closed world semantics according to which a typed feature term denotes the set of minimal feature structures which it subsumes.[2] To illustrate, let us make the declarations in (5)–(6).

(5) $\quad$ a = *a1* ∨ *a2*

(6) a. $\; {}_{a}\begin{bmatrix} \text{F}: & f \end{bmatrix}$

b. $\; {}_{a1}\begin{bmatrix} \text{G}: & g \end{bmatrix}$

c. $\; {}_{a2}\begin{bmatrix} \text{H}: & h \end{bmatrix}$

Given these constraints, a feature structure can be of type a only if it is either an *a1* or an *a2*. Moreover, any object of type *a1* or *a2* must also satisfy the constraint associated with *a*. As a consequence, we take (6a) to be semantically equivalent to the set of specializations in (7).

(7)
$$ \{ {}_{a1}\begin{bmatrix} \text{G}: & g \\ \text{F}: & f \end{bmatrix}, {}_{a2}\begin{bmatrix} \text{H}: & h \\ \text{F}: & f \end{bmatrix} \}. $$

[2] Following Pollard and Sag (1993), the feature structures in this set should be totally well-typed (i.e. for every node in the structure, all and only the features appropriated for that node are present) and sort-resolved (i.e. every node in the structure is assigned a minimal type).

3 Sierra Miwok

Goldsmith (1990) uses data involving Sierra Miwok verb stems to illustrate morphologically determined alternations in skeletal structure. He discusses three of the four classes[3] of stem, where the division into classes depends primarily on the syllable structure of the basic form, which is the form used for the present tense.

As shown in (8), each type has forms other than the basic one, depending on the morphological or grammatical context; these additional forms are called second, third and fourth stems.

(8)

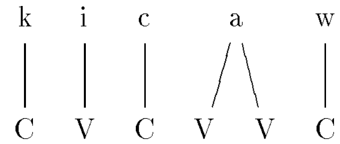
Gloss	Basic stem	Second stem	Third stem	Fourth stem
Class I				
bleed	kicaaw	kicaww	kiccaw	kicwa
jump	tuyaaŋ	tuyaŋŋ	tuyyaŋ	tuyŋa
take	patɨɨt	patɨtt	pattɨt	pattɨ
roll	huteel	hutell	huttel	hutle
Class II				
quit	celku	celukk	celluk	celku
go home	woʔlu	woʔull	woʔʔul	woʔlu
catch up with	nakpa	nakapp	nakkap	nakpa
spear	wimki	wimikk	wimmik	wimki
Class III				
bury	hamme	hameʔʔ	hammeʔ	hamʔe
dive	ʔuppi	ʔupiʔʔ	ʔuppiʔ	ʔupʔi
speak	liwwa	liwaʔʔ	liwwaʔ	liwʔa
sing	mɨlli	mɨliʔʔ	mɨlliʔ	mɨlʔi

From the perspective of generative phonology, such data require a battery of morphologically conditioned phonological rules. Assuming that the basic forms provide underlying representations, consider the derivation of the non-basic stems from **kicaaw**. First, the long vowel must be shortened. Then, for the second stem, the final consonant is geminated; in the third stem, the medial consonant must be geminated; and the fourth stem requires metathesis of the final vowel and consonant. Unfortunately, these rules are not enough, for they do not generalise to the other classes (assuming still that the basic stems provide the underlying forms). For example, the rule which derives the second stem **celukk** from **celku** by gemination of the final consonant must be preceded by yet another rule which metathesises the final consonant and vowel of the basic stem.

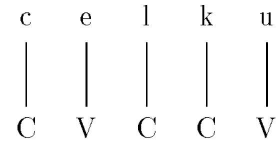
Goldsmith (1990) shows convincingly that a more insightful approach is possible within an autosegmental model. The key insight is that each stem form provides a characteristic skeleton. He initially assigns the three classes the following autosegmental representations:

[3] I use the term 'class' in preference to Goldsmith's 'type' in order to avoid possible confusion with the types of the grammar formalism.

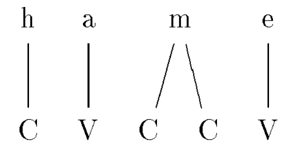
(9) a. Class I



b. Class II



c. Class III



He then argues:

> The lexical representation of verbs will look much like their 'basic stem' forms. To form the second, third and fourth stems, the skeletal strings in [(10a, b, c)] are used to replace the lexical skeletal tier of the verb stem. (Goldsmith 1990:87)

(10) a. Second stem CVCV(C)C

b. Third stem CV(C)CVC

c. Fourth stem CVCCV

Note that Goldsmith here uses parentheses to signal the first skeleton slot of a geminate consonant; I shall return later to this device.

In the following analysis, I shall adopt Goldsmith's central claim that phonological exponent of the stem is a skeleton together with an appropriate anchoring. Unlike Goldsmith, I shall not attempt here to give a general encoding of association, but confine my attention to the manner in which skeletal slots are linked to the consonant and vowel melodies by the morpho-phonology.

I should note one other significant respect in which I depart from his analysis. Goldsmith assumes that the lexical representation of each Sierra Miwok verb contains a skeletal pattern, namely that of the basic form. But as long as the grammar indicates which class a particular verb belongs to, a specification of skeletal structure is unnecessary at the lexical level. Instead, we can uniformly treat all of the stem forms, including the basic one, as morphological modifications of a verb root which lacks skeletal information.

4 Encoding Phonology

A large amount of HPSG apparatus can be given a computational interpretation. The small fragment of phonology discussed below was designed to stay within the standard data structures employed by such implementations. As a result, phonological sequences are encoded not as strings, but rather as lists, delimited by angle brackets. For example, the string **kicaaw** is represented as ⟨**k i c a a w**⟩.

Consider again the skeletal structure of Class I verb stems shown above in (9). As Goldsmith (1990) points out, there is a closely related representation which differs only in that the CV information is split across two tiers (and which allows a more elegant account of metathesis):

(11)

$$\begin{array}{lcccccc}
\text{consonantal melody} & \mathbf{k} & & \mathbf{c} & & & \mathbf{w} \\
 & | & & | & & & | \\
\text{skeleton} & X & X & X & X & X & X \\
 & & | & & \backslash & / & \\
\text{vowel melody} & & \mathbf{i} & & \multicolumn{2}{c}{\mathbf{a}} &
\end{array}$$

(11) can be translated into the following attribute-value expression:

(12)
$$\begin{bmatrix} \text{CON} : & \langle \boxed{1}\mathbf{k}\ \boxed{3}\mathbf{c}\ \boxed{5}\mathbf{w} \rangle \\ \text{VOW} : & \langle \boxed{2}\mathbf{i}\ \boxed{4}\mathbf{a} \rangle \\ \text{SKEL} : & \langle \boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{4}\ \boxed{5} \rangle \end{bmatrix}_{phon}$$

The coindexing shows how segments on the consonant and vowel tiers are anchored to the skeleton. For example, the anchoring of **k** to the first slot in the skeleton tier is encoded by the index $\boxed{1}$ which appears in both positions. Since association in (11) consists only of slot-filling (rather than the more general temporal interpretation discussed in Bird and Klein (1990)), the coindexing representation seems to be semantically adequate.

5 Roots and Stems

The analysis starts from the assumption that the Sierra Miwok lexicon will contain *minimally redundant* entries for the three classes of verb. Following my remarks at the end of section 3, lexical entries will consist of roots, not stems. Let us consider the root corresponding to the basic stem form **kicaaw**. We take the unpredictable information to be the consonantal and vowel melodies, the valency, the semantics, and the fact it is a Class I verb root. This is stated as (13) (where class membership is indicated by the type *v-root-I*).

(13)
$$\begin{bmatrix} \text{PHON} : & \begin{bmatrix} \text{CON} : & \langle \mathbf{k}\ \mathbf{c}\ \mathbf{w} \rangle \\ \text{VOW} : & \langle \mathbf{i}\ \mathbf{a} \rangle \end{bmatrix}_{phon} \\ \text{SYNSEM} : & \begin{bmatrix} \text{SUBCAT} : & \langle NP \rangle \\ \text{SEM} : & bleed \end{bmatrix}_{synsem} \end{bmatrix}_{v\text{-}root\text{-}I}$$

Notice that we have said nothing about how the melodies are anchored to a skeleton—this will be a task for the morphology. Moreover, this entry will inherit various properties by virtue of its type. Illustrative lexical entries belonging to Classes II and III are shown in (14).

(14) a.

$$\left[\begin{array}{ll} \text{PHON}: & \left[\begin{array}{ll} \text{CON}: & \langle \mathbf{c\ l\ k} \rangle \\ \text{VOW}: & \langle \mathbf{e\ u} \rangle \end{array}\right]_{phon} \\ \text{SYNSEM}: & \left[\begin{array}{ll} \text{SUBCAT}: & \langle NP \rangle \\ \text{SEM}: & quit \end{array}\right]_{synsem} \end{array}\right]_{v\text{-}root\text{-}II}$$

b.

$$\left[\begin{array}{ll} \text{PHON}: & \left[\begin{array}{ll} \text{CON}: & \langle \mathbf{h\ m} \rangle \\ \text{VOW}: & \langle \mathbf{a\ e} \rangle \end{array}\right]_{phon} \\ \text{SYNSEM}: & \left[\begin{array}{ll} \text{SUBCAT}: & \langle NPNP \rangle \\ \text{SEM}: & bury \end{array}\right]_{synsem} \end{array}\right]_{v\text{-}root\text{-}III}$$

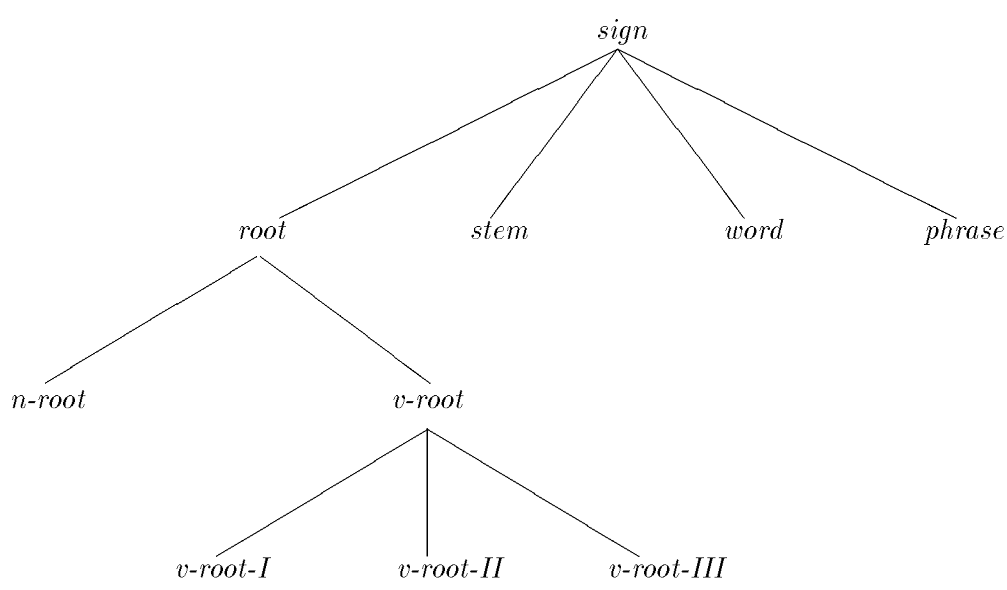
The three classes of verb root share at least one important property, namely that they are all verbs. This is expressed in the next two statements:

(15) $v\text{-}root = v\text{-}root\text{-}I \vee v\text{-}root\text{-}II \vee v\text{-}root\text{-}III$

(16) $$\left[\text{SYNSEM}: \left[\text{CAT}: verb\right]\right]_{v\text{-}root}$$

We will also assume, for generality, that roots can be divided into at least *v-root*s and *n-root*s. The hierarchy we have constructed so far looks as follows:

(17)



The next step is to show how a *v-root-I* undergoes morphological modification to become a basic verb stem; that is, a form with skeletal structure. There are of course a number of routes one might adopt for morphological modification, and in the next section I will briefly consider two of them.

5.1 Morphological Modification

Given the syntactic framework of HPSG, it seems tempting to handle morphological modification in a manner analogous to syntactic phrases. That is, morphological heads would be analysed as functors which are subcategorised for arguments of the appropriate type, and morphemes would combine in a Word Syntax scheme.[4] Simplifying drastically, such an approach would analyse the English third person singular present suffix -**s** in the manner shown in (18), assuming that affixes are taken to be heads.

(18)
$$_{\textit{affix}}\begin{bmatrix} \textsc{phon}: & \langle \mathbf{s} \rangle \\ \textsc{synsem}: & \begin{bmatrix} \textsc{subcat}: & \langle \textit{v-root} \rangle \end{bmatrix} \end{bmatrix}$$

By adding appropriately modified versions of the Head Feature Principle, Subcategorization Principle and linear order statements, such a functor would combine with a verb root to yield a tree-structured sign for the word form **walks**.

(19)
$$_{\textit{verb}}\begin{bmatrix} \textsc{phon}: & \langle \mathbf{w\text{ɔ}{:}ks} \rangle \\ \textsc{dtrs}: & \left\langle {}_{\textit{v-root}}\begin{bmatrix} \textsc{phon}: & \langle \mathbf{w\text{ɔ}{:}k} \rangle \end{bmatrix} {}_{\textit{affix}}\begin{bmatrix} \textsc{phon}: & \langle \mathbf{s} \rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

However, there seems to be no compelling reason why affixes should be admitted as grammatical objects in their own right, once we have the ability to encode specialization by means of subtyping. Inflected forms can be treated as types of *word*s which have a specification for the attribute ROOT.[5] (20) illustrates how we could declare the third person singular (*3ps*) as a subtype of *verb* in English (where $\circ$ is the append operation on lists—standing in for concatenation on strings):

(20)
$$_{\textit{3ps}}\begin{bmatrix} \textsc{phon}: & \boxed{1} \circ \langle \mathbf{s} \rangle \\ \textsc{root}: & {}_{\textit{v-root}}\begin{bmatrix} \textsc{phon}: & \boxed{1} \end{bmatrix} \end{bmatrix}$$

Although (20) contains a slot for a *v-root* (still unspecified), there is no affix as such which encapsulates the third person inflection. Instead, (20) fits better into a Word and Paradigm notion of morphology (cf. Matthews (1974)), where members of a paradigm consist of a cluster of subtypes.

Given a closed world semantics for types, the interpretation of the subterm of type *v-root* in (20) is the set of all the minimal feature structures it subsumes. Suppose, for example, that our lexicon contained only two instances of *v-root*s, namely **walk** and **meet**. Then (20) would evaluate to exactly two fully specified word forms, where *v-root* was expanded to the signs for **walk** and **meet** respectively. (21) illustrates the case for the first of these options.

[4] See Krieger and Nerbonne (1991) for an analysis of this sort within HPSG.

[5] See Riehemann (1992) for a detailed working out of a similar idea for German **-bar** derivation.

(21)
$$\left[\begin{array}{ll} \textsc{phon}: & \boxed{1} \circ \langle \mathbf{s} \rangle \\ \textsc{root}: & \left[\begin{array}{ll} \textsc{phon}: & \boxed{1} \langle \mathbf{w\text{ɔ}{:}k} \rangle \\ \textsc{synsem}: & \left[\textsc{cat}: \ \mathit{verb} \right] \end{array}\right]_{\mathit{v\text{-}root}} \end{array}\right]_{\mathit{3ps}}$$

The second, paradigm-oriented approach is the one which I shall adopt in the analysis in the following sections.

5.2 Morphological Exponency

Let us return to the task we were concerned with above, namely the morpho-phonological mapping from verb roots into verb stems. First we classify stems:

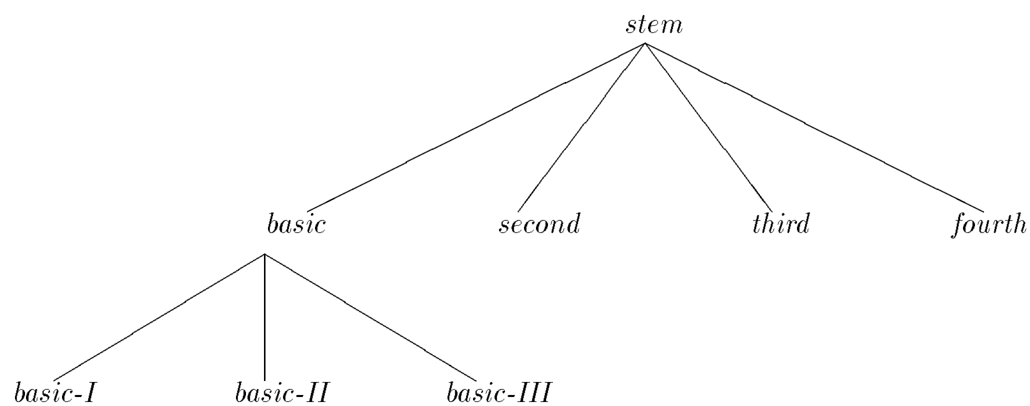
(22) $\mathit{stem} = \mathit{basic} \vee \mathit{second} \vee \mathit{third} \vee \mathit{fourth}$

Then we further subclassify basic stems:[6]

(23) $\mathit{basic} = \mathit{basic\text{-}I} \vee \mathit{basic\text{-}II} \vee \mathit{basic\text{-}III}$

Graphically, then, we have the following inflectional hierarchy of stem forms.

(24)



The next question to address is how to characterise *stems* in general (i.e. as opposed to their subtypes). For our Sierra Miwok data, we just require that a *stem* inherits the PHON and SYNSEM values of its root.

(25)
$$\left[\begin{array}{ll} \textsc{phon}: & \boxed{1} \\ \textsc{synsem}: & \boxed{2} \\ \textsc{root}: & \left[\begin{array}{ll} \textsc{phon}: & \boxed{1} \\ \textsc{synsem}: & \boxed{2} \end{array}\right] \end{array}\right]_{\mathit{stem}}$$

[6]The node *basic* in this hierarchy is only motivated by concerns of conceptual clarity; however, it is likely that a more detailed analysis of Sierra Miwok would identify morpho-syntactic properties associated with just the basic forms, and these would be stated as contraints on *basic*.

Each of the subtypes of *stem* will specialize the value of the top-level attribute PHON in (25) above. Given the coindexing of PHON and ROOT|PHON which we declared for *stem*, this will have the consequence that the ROOT's PHON value will become further instantiated. In particular, the inflected form will add a specification of skeletal anchoring which was left undetermined in the lexical entry of roots like (13).

How, then, should the inflected forms be defined so as to derive the correct results? Take, for example, the second stem form corresponding to the root for **kicaaw**. What we need is a value of *second-phon* (i.e. the value of PHON in *second*) which meets the following constraint:

(26)
$$\left[\begin{array}{ll}\text{CON}: & \langle \mathbf{k\ c\ w}\rangle \\ \text{VOW}: & \langle \mathbf{i\ a}\rangle\end{array}\right]_{phon} \wedge \textit{second-phon} = \left[\begin{array}{ll}\text{CON}: & \langle \boxed{1}\mathbf{k}\ \boxed{3}\mathbf{c}\ \boxed{5}\mathbf{w}\rangle \\ \text{VOW}: & \langle \boxed{2}\mathbf{i}\ \boxed{4}\mathbf{a}\rangle \\ \text{SKEL}: & \langle \boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{5}\rangle\end{array}\right]_{phon}$$

This result can indeed be obtained if the type *second* is declared as kind of STEM with the following value for its PHON attribute (where as an aid to readability, the numerical tags are supplemented with a C or V to indicate the type of value involved):

(27)
$$\left[\begin{array}{ll}\text{PHON}: & \left[\begin{array}{ll}\text{CON}: & \langle \boxed{1}C\ \boxed{3}C\ \boxed{5}C\rangle \\ \text{VOW}: & \langle \boxed{2}V\ \boxed{4}V\rangle \\ \text{SKEL}: & \langle \boxed{1}C\ \boxed{2}V\ \boxed{3}C\ \boxed{4}V\ \boxed{4}V\ \boxed{5}C\rangle\end{array}\right]_{phon}\end{array}\right]_{second}$$

To see this, let us suppose that the value of ROOT in (25) is instantiated to (13), as shown in (28).

(28)
$$\left[\begin{array}{ll}\text{PHON}: & \boxed{1} \\ \text{SYNSEM}: & \boxed{2} \\ \text{ROOT}: & \left[\begin{array}{ll}\text{PHON}: & \boxed{1}\left[\begin{array}{ll}\text{CON}: & \langle \mathbf{k\ c\ w}\rangle \\ \text{VOW}: & \langle \mathbf{i\ a}\rangle\end{array}\right] \\ \text{SYNSEM}: & \boxed{2}\left[\begin{array}{ll}\text{SUBCAT}: & \langle NP\rangle \\ \text{SEM}: & \textit{bleed}\end{array}\right]\end{array}\right]_{v\text{-}root\text{-}I}\end{array}\right]_{stem}$$

This instantiation of *stem* can now be specialized to the subtype *second*. This has the effect of interpreting the value of the top-level attribute PHON as the conjunction in (26), since PHON must satisfy both the constraints coming from the ROOT's PHON and the constraint given for *second* in (27).

Exactly parallel considerations govern the declarations for the third and fourth stem forms:

(29) a.

$$ {}_{\mathit{third}}\left[\begin{array}{ll} \mathrm{PHON}: & {}_{\mathit{phon}}\left[\begin{array}{ll} \mathrm{CON}: & \langle \boxed{1}C\ \boxed{3}C\ \boxed{5}C\rangle \\ \mathrm{VOW}: & \langle \boxed{2}V\ \boxed{4}V\rangle \\ \mathrm{SKEL}: & \langle \boxed{1}C\ \boxed{2}V\ \boxed{3}C\ \boxed{3}C\ \boxed{4}V\ \boxed{5}C\rangle \end{array}\right] \end{array}\right] $$

b.

$$ {}_{\mathit{fourth}}\left[\begin{array}{ll} \mathrm{PHON}: & {}_{\mathit{phon}}\left[\begin{array}{ll} \mathrm{CON}: & \langle \boxed{1}C\ \boxed{3}C\boxed{5}C\rangle \\ \mathrm{VOW}: & \langle \boxed{2}V\ \boxed{4}V\rangle \\ \mathrm{SKEL}: & \langle \boxed{1}C\ \boxed{2}V\ \boxed{3}C\ \boxed{5}C\ \boxed{4}V\ \rangle \end{array}\right] \end{array}\right] $$

These two forms are illustrated with analyses of **celluk** and **celku**, respectively.

(30)

$$ {}_{\mathit{third}}\left[\begin{array}{ll} \mathrm{PHON}: & \boxed{6}\left[\begin{array}{ll} \mathrm{CON}: & \langle \boxed{1}\mathbf{c}\ \boxed{3}\mathbf{l}\ \boxed{5}\mathbf{k}\rangle \\ \mathrm{VOW}: & \langle \boxed{2}\mathbf{e}\ \boxed{4}\mathbf{u}\rangle \\ \mathrm{SKEL}: & \langle \boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \rangle \end{array}\right] \\ \mathrm{SYNSEM}: & \boxed{7}\left[\begin{array}{ll} \mathrm{CAT}: & \mathit{verb} \\ \mathrm{SUBCAT}: & \langle \mathit{NP}\rangle \\ \mathrm{SEM}: & \mathit{quit} \end{array}\right] \\ \mathrm{ROOT}: & {}_{\mathit{v\text{-}root\text{-}II}}\left[\begin{array}{ll} \mathrm{PHON}: & \boxed{6} \\ \mathrm{SYNSEM}: & \boxed{7} \end{array}\right] \end{array}\right] $$

(31)

$$ {}_{\mathit{fourth}}\left[\begin{array}{ll} \mathrm{PHON}: & \boxed{6}\left[\begin{array}{ll} \mathrm{CON}: & \langle \boxed{1}\mathbf{c}\ \boxed{3}\mathbf{l}\ \boxed{5}\mathbf{k}\rangle \\ \mathrm{VOW}: & \langle \boxed{2}\mathbf{e}\ \boxed{4}\mathbf{u}\rangle \\ \mathrm{SKEL}: & \langle \boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{5}\ \boxed{4}\rangle \end{array}\right] \\ \mathrm{SYNSEM}: & \boxed{7}\left[\begin{array}{ll} \mathrm{CAT}: & \mathit{verb} \\ \mathrm{SUBCAT}: & \langle \mathit{NP}\rangle \\ \mathrm{SEM}: & \mathit{quit} \end{array}\right] \\ \mathrm{ROOT}: & {}_{\mathit{v\text{-}root\text{-}II}}\left[\begin{array}{ll} \mathrm{PHON}: & \boxed{6} \\ \mathrm{SYNSEM}: & \boxed{7} \end{array}\right] \end{array}\right] $$

The situation with basic stem forms is slightly more complex. As we have seen, the three root classes have different exponents as basic stems. The first step in capturing this variance was already taken earlier when I introduced three subtypes of *basic*. Each of these receives a declaration analogous to those already used for the non-basic forms, encoding the patterns of skeletal anchoring associated with the three classes of basic stem. There is, however, one addition: namely, we specify which subtype of *v-root* needs to occur as the value of ROOT. This ensures that each class of root is correlated with the appropriate skeleton.

(32) a.

$$\left[\begin{array}{ll} \text{PHON}: & \left[\begin{array}{ll} \text{CON}: & \langle \boxed{1}C\ \boxed{3}C\ \boxed{5}C\rangle \\ \text{VOW}: & \langle \boxed{2}V\ \boxed{4}V\rangle \\ \text{SKEL}: & \langle \boxed{1}C\ \boxed{2}V\ \boxed{3}C\ \boxed{4}V\ \boxed{4}V\ \boxed{5}C\rangle \end{array}\right]_{phon} \\ \text{ROOT}: & \textit{v-root-I} \end{array}\right]_{basic\text{-}I}$$

b.

$$\left[\begin{array}{ll} \text{PHON}: & \left[\begin{array}{ll} \text{CON}: & \langle \boxed{1}C\ \boxed{3}C\ \boxed{5}C\rangle \\ \text{VOW}: & \langle \boxed{2}V\ \boxed{4}V\rangle \\ \text{SKEL}: & \langle \boxed{1}C\ \boxed{2}V\ \boxed{3}C\ \boxed{4}V\ \boxed{5}C\rangle \end{array}\right]_{phon} \\ \text{ROOT}: & \textit{v-root-II} \end{array}\right]_{basic\text{-}II}$$

c.

$$\left[\begin{array}{ll} \text{PHON}: & \left[\begin{array}{ll} \text{CON}: & \langle \boxed{1}C\ \boxed{3}C\rangle \\ \text{VOW}: & \langle \boxed{2}V\ \boxed{4}V\rangle \\ \text{SKEL}: & \langle \boxed{1}C\ \boxed{2}V\ \boxed{3}C\ \boxed{3}C\ \boxed{4}V\ \rangle \end{array}\right]_{phon} \\ \text{ROOT}: & \textit{v-root-III} \end{array}\right]_{basic\text{-}III}$$

(33) illustrates an analysis of **kicaaw** obtained by specializing *stem* to *basic-I*.

(33)

$$\left[\begin{array}{ll} \text{PHON}: & \boxed{6}\left[\begin{array}{ll} \text{CON}: & \langle \boxed{1}\mathbf{k}\ \boxed{3}\mathbf{c}\ \boxed{5}\mathbf{w}\rangle \\ \text{VOW}: & \langle \boxed{2}\mathbf{i}\ \boxed{4}\mathbf{a}\rangle \\ \text{SKEL}: & \langle \boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{4}\ \boxed{5}\rangle \end{array}\right] \\ \text{SYNSEM}: & \boxed{7}\left[\begin{array}{ll} \text{CAT}: & \textit{verb} \\ \text{SUBCAT}: & \langle NP\rangle \\ \text{SEM}: & \textit{bleed} \end{array}\right] \\ \text{ROOT}: & \left[\begin{array}{ll} \text{PHON}: & \boxed{6} \\ \text{SYNSEM}: & \boxed{7} \end{array}\right]_{v\text{-}root\text{-}I} \end{array}\right]_{basic\text{-}I}$$

Exactly the same mechanisms will produce the basic stems for the other two classes of verb root.

6 Further Issues

Inspection of the Class III verb forms in (8) suggests that they only have two lexically specified consonants. The glottal stop which appears in the non-basic forms (e.g. **hame??**, parallel to **celukk**) is interpreted by Goldsmith as a default consonant, inserted into all skeletal C-positions which have not otherwise been associated. One straightforward way of describing the data would be stipulate that the tail of the CON attribute of Class III forms is an optional glottal stop. For example, let us declare the following list types (where *elist* is the type of empty lists):

(34) a. *opt-default* = *elist* $\vee$ *default*

b. $default\langle ? \rangle$

Suppose, in addition, we replace the lexical entry (14) by the following:

$$(35)\quad \begin{bmatrix} \text{PHON}: & \begin{bmatrix} \text{CON}: & \langle \mathbf{h}\ \mathbf{m}\rangle \circ \textit{opt-default} \\ \text{VOW}: & \langle \mathbf{a}\ \mathbf{e}\rangle \end{bmatrix}_{phon} \\ \text{SYNSEM}: & \begin{bmatrix} \text{SUBCAT}: & \langle \textit{NPNP}\rangle \\ \text{SEM}: & \textit{bury} \end{bmatrix}_{synsem} \end{bmatrix}_{v\text{-}root\text{-}III}$$

Now there are two cases to consider. If (35) occurs within *basic-III* (cf. (32c) , then $\langle \mathbf{h}\ \mathbf{m}\rangle \circ$ *opt-default* will unify with the two element list $\langle \boxed{1}C\ \boxed{3}C\rangle$ provided for the top-level PHON|CON value, since *opt-default* can be instantiated as the empty list. If however the root occurs in a non-basic form, whose CON values supply three C-positions, then *opt-default* will be specialized to the singleton list containing ?, as required.

Although this is a solution of sorts, it would clearly be preferable to have an account which treated ? as a default consonant throughout the phonology, not just for a particular set of verb forms. I suspect that a more general solution to the problem depends on more general account of association. Bird and Klein (1993) show how French syllabification can be encoded as a recursive constraint on well-formed phrases, and it seems likely that a similar technique would be work for association. However, I leave this as a task for future research.

Let us turn to another observation about the phonology of the verb forms. So far, I have failed to capture the fact that all the skeletons involved have an initial CVC sequence. To what extent this is an significant generalization is unclear from the data under consideration. Nevertheless, it is interesting to note that, if we wish, we can make this a general phonological property of stems. First, we define the following constraint (where $\top$ is the maximally unspecified type):

$$(36)\quad \textit{Disyllabic} \equiv \begin{bmatrix} \text{PHON}: & \begin{bmatrix} \text{CON}: & \langle \boxed{1}\ \boxed{3}\rangle \circ \textit{list} \\ \text{VOW}: & \langle \boxed{2}\ \top\rangle \\ \text{SKEL}: & \langle \boxed{1}\ \boxed{2}\ \boxed{3}\rangle \end{bmatrix} \end{bmatrix}$$

To ensure that every stem possesses this property, we assert the constraint

(37) *stem* $\wedge$ *Disyllabic*.

And finally, we modify all the skeletal templates so that they omit the anchoring information contained in *Disyllabic*. As an illustration, here is the revised declaration for *second*:

$$(38)\quad \begin{bmatrix} \text{PHON}: & \begin{bmatrix} \text{CON}: & \langle \top\ \top\ \boxed{2}C\rangle \\ \text{VOW}: & \langle \top\ \boxed{1}V\rangle \\ \text{SKEL}: & \langle \top\ \top\ \top\ \boxed{1}V\ \boxed{2}V\ \boxed{2}C\rangle \end{bmatrix}_{phon} \end{bmatrix}_{second}$$

7 Concluding Remarks

As I indicated above, it would be desirable to have a less lexically-specific account of association, and in this respect Goldsmith's framework has the advantage. Nevertheless, it is worth noting that Goldsmith is compelled to introduce additional apparatus to account for the insertion of the default glottal stop and, more worryingly, for geminate consonants. (The first C-position of a geminate is ignored by the Association Convention, and an auxiliary rule associates it to the segment which was associated with the next C-position.) The coindexing account adopted here (and also advocated by Scobbie 1991) seems clearly preferable.

Moreover, I have tried to be much more explicit than Goldsmith about how the morphology intervenes in morphologically conditioned phonological processes. In the course of so doing, I have shown how exploiting the general framework of type inheritance can eliminate appeals to morphological affixes which lack any segmental content.

The set of constraints described in this paper can be interpreted, given a suitable constraint resolution mechanism, as defining inputs for either parsing or generating (cf. Zajac 1992). Consider, for example, the two feature terms in (39).

(39) a. $\left[\textsc{phon} : \left[\textsc{skel} : \langle \mathbf{k\ i\ c\ a\ a\ w} \rangle\right]\right]$

b. $_{\textit{basic-I}}\left[\textsc{synsem} : \left[\textsc{sem} : \textit{bleed}\right]\right]$

(39a) can be taken as a string to be parsed, and (39b) as semantic/morphological input to a generator. Both evaluate to the same feature structure, namely that represented in (33).

References

Bird, S. (1990) *Constraint-Based Phonology.* PhD thesis, Centre for Cognitive Science, University of Edinburgh.

Bird, S. and E. Klein (1990) Phonological events. *Journal of Linguistics* **26**, 33–56.

Bird, S. and E. Klein (1993) Enriching HPSG phonology. Research Paper EUCCS/RP-56, Centre for Cognitive Science, University of Edinburgh.

Carpenter, B. (1992) *The Logic of Typed Feature Structures: With Applications to Unifications Grammars, Logic Programs and Constraint Resolution.* Cambridge: Cambridge University Press.

Emele, M. C. and R. Zajac (1990) Typed Unification Grammars. In H. Karlgren, ed., *COLING-90: Papers Presented to the 13th International Conference on Computational Linguistics*, Vol. 3, pp. 293–298. University of Helsinki.

Gibbon, D. (1987) Finite State Processing of Tone Systems. In *Proceedings of the 3rd European Meeting of the Association for Computational Linguistics*, pp. 291–297.

Goldsmith, J. (1990) *Autosegmental and Metrical Phonology.* Oxford: Basil Blackwell.

Krieger, H-U. and J. Nerbonne (1991) Feature-based inheritance networks for computational lexicons. Research Report RR-91-31, DFKI, Saarbrücken.

Manandhar, S. K. (1993) *Relational Extensions to Feature Logic: Applications to Constraint Based Grammars.* PhD thesis, Department of Artificial Intelligence, University of Edinburgh.

Matthews, P. H. (1974) *Morphology: An Introduction to the Theory of Word Structure.* Cambridge: Cambridge University Press.

Pollard, C. and I. A. Sag (1987) *Information-Based Syntax and Semantics*, Vol. 1: *Fundamentals.* Stanford, Ca.: Center for the Study of Language and Information.

Pollard, C. and I. A. Sag (1993) *Head-Driven Phrase Structure Grammar.* Chicago, Ill.: University of Chicago Press and CSLI Publications.

Riehemann, S. (1992) Word formation in lexical type hierarchies: A case study of *bar*-adjectives in German. Master's thesis, Department of Linguistics, University of Tübingen.

Scobbie, J. M. (1991) *Attribute Value Phonology.* PhD thesis, Centre for Cognitive Science, University of Edinburgh.

Zajac, R. (1992) Inheritance and constraint-based grammar formalisms. *Computational Linguistics* **18**(2), 159–182.

Appendix

The grammar constraints described in this paper have been implemented in the Stuttgart TFS grammar development system (Emele and Zajac 1990; Zajac 1992). The listing is given below.

```
:KB Sierra-Miwok.

:ATTRIBUTE-ORDER PHON,
                 SYNSEM, CAT, HEAD, CONT,
                         FIRST, REST, 1, 2, 3, CON, VOW, SKEL,
                 ROOT.

;;; Definitions from Emele's HPSG grammars
;;; -------------------------------------

:NIL elist.     ;; type symbol for the empty list.
:CONS nelist.   ;; type symbol for the non-empty list.
:CAR FIRST.     ;; feature symbol for the head of a non-empty  list.
:CDR REST.      ;; feature symbol for the tail of a non-empty  list.

:TOP bottom.    ;; dual ordering
:BOTTOM top.

list = elist | nelist.

nelist[FIRST: bottom, REST: list].

;;; Maps fixed arity to features
append(#X,#Y,#Z) := append[1:#X, 2:#Y, 3:#Z].

append(<>,#Ls,#Ls).
append(<#X.#Xs>,#Ys,<#X.#Zs>) :- append(#Xs,#Ys,#Zs).

;;;Signs
;;;-----

sign = root | stem | word | phrase.

sign[PHON: phon,
     SYNSEM: synsem].

synsem[CAT:  cat,
       SUBCAT: list,
       CONT: cont].

cont[RELN: bottom].
```

```
cat = noun | verb | adj | prep.

;;;Some phonology
;;;--------------


phon[CON: list,
     VOW: list,
     SKEL: list].


Disyll := [PHON: [CON: <#C1 #C2 . list>,
                  VOW:  <#V1 bottom>,
                  SKEL: <#C1 #V1 #C2 . nelist>]].


;;;Morpho-Syntax
;;;--------------

stem = basic | second | third | fourth.

basic= basic-I | basic-II | basic-III.

root = v-root | n-root.

v-root = v-root-I | v-root-II | v-root-III.

v-root[SYNSEM: [CAT: verb]].

stem[   PHON: #p,
        SYNSEM: #s,
        ROOT: v-root[   PHON: #p,
                                SYNSEM:#s]] & Disyll.


;;;Inflections
;;;-----------

;;; Basic Stem Forms


basic-I[PHON: phon[      CON: [REST:[REST: <#C3>]],
                         VOW: [REST: <#V2>],
                         SKEL: [REST:[REST:[REST: <#V2 #V2 #C3>]]]],
        ROOT: v-root-I].
```

```
basic-II[PHON: phon[    CON: [REST:[REST: <#C3>]],
                        VOW: [REST: <#V2>],
                        SKEL: [REST:[REST:[REST: <#V2 #C3>]]]],
         ROOT: v-root-II].

basic-III[PHON: phon[CON: [REST:[REST: elist]],
                        VOW: [REST: <#V2>],
                        SKEL: [REST:[REST: <#C2 #C2 #V2>]]],
          ROOT: v-root-III].

;;; Non-Basic Stem Forms


second[PHON: phon[CON:  [REST:[REST: <#C3>]],
                  VOW:  [REST: <#V2>],
                  SKEL: [REST:[REST:[REST: <#V2 #C3 #C3>]]]]].

third[PHON: phon[CON:  [REST:[REST: <#C3>]],
                  VOW:  [REST: <#V2>],
                  SKEL: [REST:[REST: <#C2 #C2 #V2 #C3>]]]].

fourth[PHON: phon[CON:  [REST:[REST: <#C3>]],
                  VOW:  [REST: <#V2>],
                  SKEL: [REST:[REST: <#C2 #C3 #V2>]]]].



;;; Roots
;;;-----------


v-root-I[PHON: phon[CON:  <k c w>,
                    VOW:  <i a>],
         SYNSEM: [      SUBCAT: <NP>,
                        CONT: [ RELN: bleed]]].

v-root-II[PHON: phon[CON:  <c l k>,
                     VOW:  <e u>],
          SYNSEM: [     SUBCAT: <NP>,
                        CONT:[  RELN: quit]]].

v-root-III[PHON: phon[CON:  <h m.list>,
                      VOW:  <a e>],
           SYNSEM: [    SUBCAT: <NP NP>,
                        CONT:[ RELN: bury]]].
```