# Towards morphology and agreement in Embodied Construction Grammar

Ben Bergen
bergen@hawaii.edu
*March 18, 2003*

## 0. General introduction

Embodied Construction Grammar (ECG) is a theory of linguistic knowledge and its use, as well as a formalism for generating analysis within that theory. Until present, ECG has predominantly been applied to syntactic and semantic phenomena (Bergen and Chang, To Appear, Chang et al. 2002), leaving more or less implicit the role and representation of morphology in the theory. The present document responds to this lack by outlining the use of ECG for certain types of morphological analysis by linguists and for Natural Language Understanding. In particular, it focuses on the representation of agreement features, including case, number, and person (Section 1), representing inflectional morphemes as meaningful or as meaningless (Section 2), and the treatment of allomorphy (Section 3), before turning in detail to the different ways morphological complexity can be represented in and ECG morphology (Section 4). Throughout, the paper assumes familiarity with the ECG paradigm and notation, which can be gained from a general introduction, like Bergen and Chang (To Appear) or Chang et al. (2002).

The main benefit of using ECG for morphological analysis is its representational flexibility. While some morphological theories are bound to a single type of morphological operation or relation, ECG allows morphologically related forms to be linked in any one of a number of ways - one may inherit the other or their may inherit from a common parent, one may be a constituent of the other, or there may be systematic correlations between them expressed as maps. The reason for seeing representational flexibility as an advantage is that the more options a morphological formalism allows, the more likely a morphological theory expressed in terms of that formalism is to get the phenomena it is attempting to capture right. Language users seem to recruit all sorts of knowledge when using language, including memorization, inheritance, constituency, and mappings, so a formalism should allow for any of these mechanisms to be incorporated into a theory of what those language users are doing. Any theory that does not allow all these possibilities, the argument goes, will necessarily miss some crucial aspect of human linguistic knowledge, even though it might be a mechanistically simpler theory. More discussion of this argument can be found in Section 5, below.

In sum, in the interest of a maximal expression, the ECG formalism allows morphological complexity to be expressed in several ways, some or all of which may play a part in a given morphological theory. Those mechanisms are the following four:

1. Sometimes, "morphologically complex" words are stored as wholes in the mind of the language user, often as a result of high frequency of use. Such words can be represented without internal morphological structure through **stored representations** that have no internal complexity.
   e.g. a word like *goes* might be a lexical construction with no internal constituents – it might be specified as having the form *goes* and being an inflected present-tense third person singular verb.

2. Other times, semantic and formal morphological characteristics of a class are inherited from an abstract representation for that class through **inheritance**. Here, the inflected or derived forms are simply more fully specified subcases of more general lexical constructions.
   e.g. a word like *jumps* might inherit from two more abstract constructions – a generalized Jump construction that pairs the form *jump* with the Jump schema, and a third person Singular-Present-Tense-Verb construction, which has no specific phonological form but has

in its contructional and meaning blocks specifications that this is an inflected present-tense third person singular verb.

3. Some cases of morphological complexity can be straightforwardly seen as the product of morphological constructional **constituency**, where lexical constructions have multiple morphological constituents (such as a noun and a plural suffix).
   e.g. a word like *jumps* might have two constructional constituents, a Jump and a Plural-Suffix, and combines these in the appropriate order, applying the appropriate bindings to the semantic components of each.

4. Finally, some morphological relations are best captured through **maps**, which take the semantic and formal structures of a given word and specify the corresponding semantic and formal structures for morphologically related words.
   e.g. related words like *give* and *gave* might be related through a map that takes a present tense verb with an /I/ in its form pole and relates this to a past tense verb that has an /ey/ at the corresponding place in its form pole.

Each of these four mechanisms may be more or less useful for analyzing a given phenomenon. For example, "morphologically complex" forms that are stored as wholes due to their frequency or other reasons (like *goes*, perhaps) and suppletive forms (like *went*) can be represented by stored representations (mechanism 1). Phonaesthemes (like *gl-* meaning 'light' and 'vision') and cran- morphs (like *cran-* might be best dealt with by inheritance (2). Truly compositional morphology (like the English regular plural) might best be dealt with using constituency (3), and ablaut (as in *fly-flew*) might be easiest to deal with using maps (4). We will say more about these different mechanisms in Section 4, but not before we discuss agreement features in the next section.

## 1. Agreement - case study of English present tense inflection

One major function of morphology is agreement, and this first section is concerned with fundamental issues surrounding this phenomenon. Four things are necessary in and ECG morphology for representing agreement.

1. Representations of constructional agreement roles on each constituent construction
2. A way of expressing semantic properties that are relevant to agreement
3. A mechanism for automatically propagating those semantic features into the constructional agreement features where possible.
4. Constructions that unify or match the constructional agreement features of constituent constructions

We will begin with a superficially simple example to walk through the four mechanisms necessary for agreement. The basic phenomena we want to capture are the following.

- There is a different form for the third personal singular form of regular verbs than for the other present tense indicative forms.
- The difference is that the third person singular form has an additional alveolar fricative at the end (we're not concerned for the time being with the allomorphy of the suffix).

### 1.1. Representations on agreeing constructions

Constructions that agree with others bear an indication of their agreement properties, no matter what part of speech they are. English present tense verbs agree with their (nominal) subjects, so after we outline agreement features for verbs, we will discuss nouns as well.

All inflected verb forms in English (and indeed on all agreeing constructions) have constructional roles, which are agreement categories, like person and number. These are seen as constructional and not semantic and formal features for simple reasons. While agreement features certainly have semantic correlates (case, person and number being prime examples of such semantically bound agreement categories), we must nevertheless separate the semantic function of a given class from its constructional representation since in some cases agreement is arbitrary or even contradicts semantic properties of the construction in questions. Gender agreement is a good example of this - it is hard to see what semantic different could distinguish French masculine *siege* 'seat' from feminine *chaise* 'chair' or for that matter German feminine *Brücke* 'bridge' from French masculine *pont* 'bridge'. Even more disturbing, some semantically gendered entities are assigned divergent linguistic genders, such as German neuter *Mädchen* 'girl'. So linguistic agreement features cannot be strictly identical to their semantic correlates, although they may be related in some way.

By the same token, agreement features do not appear to belong strictly to domain of linguistic form. That is, agreement features are not best seen as analogous to ordering constraints between linguistic elements or phonological schemas - the way a word is pronounced. While order and phonology are directly accessible from the linguistic input and directly implementable into a piece of linguistic output, agreement features are slightly more abstract constructional properties - they have everything to do with what kind of construction a given verb or noun is. Moreover, agreement relations are centrally relations among constructions, not among linguistic forms or linguistic meanings. For these reasons, we place agreement features in the constructional block, rather than in the form or meanings blocks of linguistic constructions.

Turning back to the agreement properties of verbs, all verbs in English inherit two constructional roles - person and number - directly or indirectly from a general Verb construction. The constructional roles of a given word have specified values. For example, in the Tosses construction below (Figure 1), the person role has the value 3$^{rd}$ and the number role has the value singular – value assignment is notated with a leftwards arrow (←) from the value to the feature.

```
construction TOSSES
    subcase of VERB
    constructional
        constraints
            person ← 3rd
            number ← singular
    form
        self.phon ← /tasIz/
    meaning
        …
```

Figure 1 - Simplified Tosses construction

We need also to represent the remainder of the present tense indicative verb forms. There are two types of representation we might chose to represent the fact that in all other number and person combinations, the form is /tas/. In fact, this is a relatively complex generalization, since it requires either a disjunction inside a construction or a proliferation of constructions, as seen below.

For the disjunction solution, we can represent all cases but the 3$^{rd}$ person plural by leaving the roles person and number unfilled, but constraining them to not be the conjunction of person ← 3$^{rd}$ and number ← singular. Thus, the inflected present tense forms of Toss, aside from the 3$^{rd}$ person singular would appear as in Figure 2.

```
construction TOSS
    subcase of VERB
    constructional
        constraints
            NOT (person ← 3rd AND number ← singular)
    form
        self.f.phon ← /tas/
    meaning
        …
```

Figure 2 - Simplified Toss construction using a disjunction solution

This is not a particularly elegant solution in that it requires some complexity inside the constructional constraints. It may however be preferable in certain cases to the proliferation-of-constructions solution. For this second solution, each number and person combination for each verb is positively assigned a value for the person and number role, which leads to a large number of very similar verbal constructions. For example, the 1st person singular and 3rd person singular might be represented as in Figures 3 and 4 below:

```
construction TOSS
    subcase of VERB
    constructional
        constraints
            person ← 1st
            number ← singular
    form
        self.f.phon ← /tas/
    meaning
        …
```

Figure 3 - Simplified 1st person singular Toss construction

```
construction TOSSES
    subcase of VERB
    constructional
        constraints
            person ← 3rd
            number ← plural
    form
        self.f.phon ← /tasIz/
    meaning
        …
```

Figure 4 - Simplified 3rd person plural Toss construction

Depending on the degree of syncretism in the paradigm under consideration, the disjunction or the proliferation solution might be preferred. In general, the more syncretism in a paradigm, that is, the more identity there is among forms that have different agreement features, the more favorable the disjunction solution appears, because this avoids significant proliferation. However, isolated cases of sycretism are more easily treated by multiplying formally identical constructions.

It should be noted that neither of these solutions is quite as bleak as they first appear, since it isn't the case that every set of morphologically related words need necessarily be represented explicitly and separately in a morphological model. Morphological paradigms can be represented abstractly, such that 3rd person singular verbs in general are represented as having some particular properties, which are inherited and specified by inheriting subcases (specific third person singular verb forms).

Not only verbs but also nouns have agreement properties marked on them. All nouns inherit from a general Referring-Expression construction, described in more detail in Bergen and Chang (To Appear) and in Section 4 below. Referring expressions are specified as having constructional agreement roles, including person, number, and case, and particular nouns like *Mary* and *I* fill in those roles as seen in Figures 5 and 6 below, respectively. (The meaning attributes of these two constructions will be discussed in the next subsection.)

```
construction MARY
    subcase of REF-EXPR
    constructional
        constraints
            person ← 3rd
            number ← singular
    form
        self_f.phon ←  /mEriy/
    meaning
        roles
            r : Mary
        constraints
            r.accessibility ← inactive
```

Figure 5 - Mary construction, with agreement information

```
construction I
    subcase of REF-EXPR
    constructional
        constraints
            person ← 1st
            number ← singular
            case ← subject
    form
        self_f.phon ← /ay/
    meaning
        constraints
            self_m ← current-space.speaker
            number ← 1
            accessibility ← active
```

Figure 6 - I construction

In more complicated cases, NPs may have multiple agreement features (see Dodge and Wright 2002). We have seen how verbs and nouns are specified for constructional agreement features. Now we will turn to the relation between these constructional features and their semantic correlates.

### 1.2. Semantic agreement features

In the meaning pole of nominal constructions, and perhaps in some inflected verbal constructions, semantic schemas that correspond to these constructional agreement features are evoked. For example, *I* refers to a singular speaker. As we saw above, the constructional person role being assigned a value of 1 captures the constructional ramifications of first personhood, and singularity is captured by the value assignment of the value singular to the number role. The semantic correlates of these features, however, are specified elsewhere.

Some semantic agreement features are predictable from the semantic categorization of the referent. For example, as seen in Figure 5 above, *Mary* has a referent r which is an instance of the Mary schema. This schema is a subcase of Person, which is a type of (singular) Entity. Thus, the semantic pole of the Mary construction is singular. Similarly, Mary is not specified as referring to either the speaker or hearer, and thus will necessarily function as a third person nominal construction. Neither the semantic person or number of Mary needs to be explicitly identified in the Mary construction.

By contrast, there are other cases where it is vital that semantic agreement features be directly marked on a given construction. For example, in the case of *I*, it is important to identify the referent as singular, so as to distinguish it from *we*. The fact that the I construction identifies its semantic pole as referring to the speaker (as seen in the meaning pole of the I construction in Figure 6, above) does not distinguish between a singular and plural first person, and although one can usually determine from context whether one or two speakers is talking, this does not adequately capture the semantic difference between *I* and *we*. Therefore, the I construction assigns a value of 1 to its semantic feature number. Additionally, and as somewhat of a digressions, the I construction also assigns its referent to be active, in terms of Lambrecht's (1994) classification of the accessibility discourse referents.

### 1.3. Percolation mechanism

Constructional agreement features and their semantic correlates are, as argued above, not unrelated. For this reason, a mechanism is required to get semantic agreement feature to percolate up to the constructional level. This is effected in ECG through an automatic assignment of the values of semantic agreement roles to their constructional counterparts. Thus, all nouns and verbs will inherit a unification constraint that specifies constructional features as unifying with their semantic counterparts. For verbs, this may be expressed as in Figure 7 below, using a subscript $_m$ to identify the number and person roles as pertaining to the meaning pole.

```
construction REF-EXPR
    constructional
        constraints
            number ← r_m.number
            person ← r_m.person
    form
        self_f.phon : Schematic-Form
    meaning
        roles
            r : Referent
```

Figure 7 - Ref-Expr construction

As mentioned above, we will occasionally find mismatches between semantic and constructional agreement features, like the case of German Mädchen ‘girl’, which is semantically female but constructionally neuter. The possibility of such mismatches implies that the assignment of semantic agreement feature values to their constructional counterparts according to the mechanism described in the previous paragraph must be overrideable. In other words, while all German nouns will inherit a constraint that their constructional gender should be filled in by the value of their semantic gender, specific nouns should be able to assign particular idiosyncratic values to their constructional gender role that override this constraint.

Cases like Mädchen also imply that the relation between semantic and constructional agreement features should be one of filling in of the constructional role rather than binding. In other words, the general constraint should not be that the semantic and constructional agreement features are bound together (↔), because this would imply that German speakers believed that girls were actually semantically neuter. Rather, the value of the semantic role is by default also assigned to the corresponding constructional feature.

In addition to assigning semantic features values to the constructional agreement features, the Ref-Expr construction states that all referring expressions have some form (that is, a form that is a subtype of the most general form there is – the Schematic-Form) and that they have a semantic role which is an instance of the Referent schema seen below.

```
schema Referent
    roles
        category
        distribution
        gender
        number
        quantifiers
        givenness
        case
        resolved-ref
```

The Referent schema has a number of roles, including category (the schema it instantiates), number (singular, plural, etc.), and case. Thus, the meaning role r of any Referring-Expression is an instance of the Referent schema, and as such, its roles like number and case can be assigned values.

There are two possible implementations of the fillers of the role case in the Referent schema. In the first implementation, the possible fillers of the case role are simply atomic values, which are related to a semantic schema or set of semantic schemas. In this interpretation, these feature values, like plural and 1st person in the Referent schema, are parameterizations of linguistically relevant schemas that the particular language encodes grammatically. In a second interpretation, a set of semantic schemas can be possible fillers of these particular roles (e.g. the SPG schema can fill the case role). In either of these interpretations, the fillers of the case role encode one aspect of thinking for speaking. There are some advantages to implementing them as atomic features, not least that they can serve to unify a set of semantic schemas that linguistic constructions refer to.

### 1.4. Matching constructions

We’ve seen how constructional agreement roles are represented on each constituent construction, how semantic properties that are relevant to agreement are expressed, and how those semantic features are automatically propagated into the constructional agreement features where possible. Now all that remains is to specify how constructions bearing agreement features are matched up.

For English present tense inflection, we need a mechanism to pair a subject that has a particular constructional agreement specification with an appropriate verb form. We'll use a simplified version of the Intransitive-Clause construction. This construction, which inherits from the general Pred-Expr construction, has two constructional constituents, one that's an instance of a Ref-Expr, and one that's a Verb. This construction will form the backbone of intransitive clauses, like *John fell*, *I breathed*, and so on.

The Instransitive-Clause construction, seen in Figure 8, performs two essential agreement functions: (1) asserting bindings between the constructional agreement features of its agreeing constituents and (2) assigning values to other agreement features. It identifies its two constituents as theme and verb and binds together their constructional person and number features (the two-headed arrows indicate binding). This binding serves the purpose of verification during the language analysis process - if the person and number features are incompatible (that is, cannot be unified), then the utterance cannot be said to be licensed by the Intransitive-Clause construction. This construction also assigns the constructional case of subject to its nominal constituent theme.

**construction** INTRANSITIVE-CLAUSE
  **subcase of** PRED-EXPR
  **constructional**
    **constituents**
      theme : REF-EXPR
      verb : VERB
    **constraints**
      verb.person ↔ theme.person
      verb.number ↔ theme.number
      theme.case ← subject
  **form**
    **constraints**
      $theme_f$ *before* verb
  **meaning**
    …

Figure 8 - Simplified Intransitive-Clause construction

In general then, phrase-structural constructions will perform the function of unifying constructional agreement features of their constituents. Some phrasal constructions may perform other agreement matching functions when unification does not suffice, such as in cases discussed in Wechsler and Zlatic (2000), for example.

We have seen for the simple case of English present-tense subject-verb agreement (1) how agreement features are expressed in the **constructional** block of verbal and nominal constructions, (2) how semantic correlates of those agreement features are expressed in the **meaning** block of those same constructions, (3) how constructional agreement features are automatically filled in on the basis of semantic agreement features where possible, and (4) how clausal constructions assign and bind together agreement features of their constituent verb and noun constructions. We will now turn to the meaning-bearing status of inflectional morphemes.

## 2. Meaningfulness of English past tense verbal inflection

In this section, we will look briefly at solutions to English past tense verbal inflection, addressing the meaningfulness of agreement markers. This will form the basis for the next section, which also looks at English past tense verbal inflection, focusing on allomorphy.

A central issue in the study of agreement is whether agreement markers, such as affixes, are meaning-bearing units or whether they are simply formal reflexes. It is not our current interest to advocate one of these positions or the other. Rather, we will stipulate that each account may be a valid hypothesis in certain cases. In line with its goal of maximum expressiveness, ECG can express each possibility.

We'll exemplify the mechanisms for the meaningful-suffix and the meaningless-suffix solutions with English regular past tense inflection. The basic idea of the two solutions is that a past tense suffix can be represented either as a separate morphological construction, in which case it is assigned some meaning, or can be part of a larger regular-past-tense-verb construction, in which case it only has meaning indirectly by dint of being part of this larger structure.

A regular past tense verb in English can be seen as composed of two constructional constituents, a root and a suffix, in which case each bears some meaning. Alternatively, it might be composed of only one constructional constituent, the root, in which case the form of the regular past tense suffix is incorporated into the larger Regular-Past-Tense-Verb construction, which in its entirety has a past tense meaning.

The first, meaningful-suffix solution is shown diagrammatically for the past tense form 'tossed' in Figure 9 below. Here, there is an abstract Regular-Past-Tense-Verb construction that has two constructional constituents. The first is an instance of the past-tense suffix construction, which pairs the form of the regular past tense suffix with its meaning. The second constructional constituent is an instance of the Verb-Root construction, of which the Toss construction is a subtype.

REGULAR-PAST-TENSE-VERB
↓ ↓
PAST-TENSE-SUFFIX TOSS IS-A→ VERB-ROOT

Figure 9 - Schematic representation of the meaningful-suffix version of Tossed

As seen in Figure 10 below, the Regular-Past-Tense-Verb construction, along with assigning aspectual (encapsulated) and tense (past) values to its semantic roles, constrains its Predication to unify with that of its root, and that of its suffix.

```
construction REGULAR-PAST-TENSE-VERB
    subcase of PAST-TENSE
    constructional
        constituents
            root : VERB-ROOT
            suffix : PAST-TENSE-SUFFIX
    form
        constraints
            root_f meets suffix_f
    meaning
        pred : Predication
        pred ↔ root.pred ↔ suffix.pred
        pred.aspect ← encapsulated
        pred.setting.time ← past
```

Figure 10 - Regular-Past-Tense-Verb construction, in a meaningful-suffix account

The root constituent of this construction must be an instance of the Verb-Root construction, which itself is a Predicating-Expression, meaning that it has a Predication in its semantic pole, as seen in Figure 11.

```
construction VERB-ROOT
    subcase of PREDICATING-EXPRESSION
    form
        self_f.phon : Schematic-Form
        meaning
            pred : Predication
```

Figure 11 - Verb-Root construction

Any construction that hopes to fill the root constituency of the REGULAR-PAST-TENSE-VERB construction must inherit from this Verb-Root construction. The Toss construction, for example, does just this, adding a semantic role that is an instance of the Toss schema, and also binding that schema to the schemas role of the Predication schema it has inherited from the Verb-Root construction, as in Figure 12.

```
construction TOSS
    subcase of VERB-ROOT
    form
        self_f.phon ← /tas/
    meaning
        pred.schemas ↔ t : Toss
```

Figure 12 - Toss construction

Recall that we are entertaining the possibility that the past tense suffix might constitute a meaning-bearing unit itself. In this case, the past tense suffix has a predication of its own (remember that the REGULAR-PAST-TENSE-VERB construction unifies the predications of its root and suffix). There are two ways this can be implemented. The REGULAR-PAST-TENSE-SUFFIX construction can specify a predication in either the roles or the evokes block of its meaning pole. The difference here is subtle – if the predication is part of the REGULAR-PAST-TENSE-SUFFIX construction, then this suffix contributes that predication to the meaning of the whole word. However, if the suffix merely evokes the predication, then the predication is to be found elsewhere, for example, in the root. We do not yet have tools for distinguishing empirically between these possibilities. These two solutions are highlighted in Figure 13 below – note that only one of these blocks – roles or evokes – will introduce a Predication.

```
construction REGULAR-PAST-TENSE-SUFFIX
    subcase of INFLECTIONAL-SUFFIX
    form
        self_f.phon ← /d/
    meaning
        roles
            pred : Predication
        evokes
            pred : Predication
        constraints
            pred.aspect ← encapsulated
            pred.setting.time ← past
```

Figure 13 - Regular-Past-Tense-Suffix construction evoking or containing a predication

The other major possibility discussed above is that an inflectional suffix bears no meaning by itself, but is simply an integrated part of a larger lexical construction. Shown diagrammatically, this solution appears as in Figure 14 below.

```
        REGULAR-PAST-TENSE-VERB
                  ↓
                TOSS    IS-A→  VERB-ROOT
```

Figure 14 - Schematic representation of the meaningless-suffix version of Tossed

This solution also uses a REGULAR-PAST-TENSE-VERB construction, but this one (seen in Figure has only one constituent, the root, in addition to a form element, the past tense suffix (which, since we are ignoring morpho-phonology until the next section, we will simply treat as /d/).

```
construction REGULAR-PAST-TENSE-VERB
    subcase of PAST-TENSE
    constructional
        constituents
            root : VERB-ROOT
    form
        elements
            suffix ← /d/
        constraints
            root_f meets suffix
    meaning
        pred : Predication
        pred ↔ root.pred
        pred.aspect ← encapsulated
        pred.setting.time ← past
```

Figure 15 - Regular-Past-Tense-Verb construction, with a single constituent

The other constructions, aside from the REGULAR-PAST-TENSE-SUFFIX construction, which is not needed, are identical in the two approaches.

Now that we have seen representations for agreement markers both with and without meaning, let us turn to the representation of morpho-phonological alternations.

## 3. Morpho-phonological alternations

The basic facts of the morpho-phonological alternations of the English regular past-tense are the following. The regular past tense suffix has three forms, [Id], and [t], and [d],. The first of these occurs after any alveolar stop, as in *spurted* and *beaded.* The second appears after any other voiceless consonant, as in *whisk*ed and *flossed*. Finally, the third, syllabic allomorph appears after any other voiced segment, as in *flowed* and *probed*.

English past tense

| *[Id] after alveolar stop* | *[t] after other voiceless Cs* | *[d] after other voiced segments* |
|---|---|---|
| spurt – spurted | whisk – whisked | flow – flowed |
| bead – beaded | floss – flossed | probe – probed |
| | reap – reaped | whiz – whizzed |
| | thatch – thatched | |

The main representational issue that arises here is similar in nature to the one we faced above in considering the English present tense: do we proliferate constructions or create one construction with internal variation? Focusing only on the form pole of the Regular-Past-Tense-Verb construction, it is possible to represent the alternation by (1) specifying a disjunction of form elements in the form block, and (2) expressing the conditions on these constraints in the constraints block of the form pole, as in Figure 16.

```
construction REGULAR-PAST-TENSE-VERB
    …
    form
        elements
            suffix : {[d] AND [t] AND [Id]}
        constraints
            root_f meets suffix
            if      (root_f.final.place ← alveolar
                    AND root_f.final.manner ← stop)
                    then suffix ← [Id]
            elseif  root_f.final.voicing ← voiceless
                    then suffix ← [t]
            else    suffix ← [d]
    …
```

Figure 16 - Simplified Regular-Past-Tense-Verb construction with conditioned constraints

To clarify the content of the solution in Figure 16, the form element suffix is constrained under **elements** to be one of [d], [t], or [Id]. Then, under **constraints**, we see that there are several conditional constraints on exactly which of these will be selected. In the first condition, where the final segment of the root has a place feature that is filled by the value alveolar, and where that same final segment's manner is filled by stop (that is, where the root ends with an alveolar stop), the suffix selected is the syllabic [Id]. However, when this condition does not hold and when the final segment of the root is voiceless, then the suffix form selecter is [t]. Finally, in all other cases, the suffix chosen is [d].

This solution has the same drawback as the one for English third person singular verbs – it introduces a great deal of complexity into construction-internal constraints. The alternative solution, which does not face this problem to as great an extent, uses multiple past tense constructions, without internal variation. For example, consider the two allomorphs shown in Figures 17 and 18 below. In these, there is a single form specified for each construction, each of which represents an allomorph of the regular past tense. In order for one or the other of these allomorph constructions to apply to combine with a given root, the constraints in its form pole must be satisfied. For example, for the [d] allomorph, the root's final segment must be voiced, but must not be and alveolar stop. This allows [d] to combine with roots with final voiced stops, as long as they are not alveolar, as in *rubbed*, and with roots with final voiced alveolars, so long as they are not stops, as in *buzzed*.

**construction** REGULAR-PAST-TENSE-VERB-D
**...**
**form**
**elements**
suffix : [d]
**constraints**
$root_f$ *meets* suffix
$root_f$.final.voicing ← voiced
NOT ($root_f$.final.place ← alveolar AND
$root_f$.final.manner ← stop)
**...**

Figure 17 - Regular-Past-Tense-Verb-D construction, for the [d] allomorph

**construction** REGULAR-PAST-TENSE-VERB-ED
**...**
**form**
**elements**
suffix ← [Id]
**constraints**
$root_f$ *meets* suffix
$root_f$.final.place ← alveolar
$root_f$.final.manner ← stop
**...**

Figure 18 - Regular-Past-Tense-Verb-Ed construction, for the [Id] allomorph

Other solutions are imaginable, such as a purely phonological solution, which inserts [I] between sequences of coronal stops in coda position, and assimilates the voicing of word-final coronal stops. Another might be to implement a formal suffix role on the root, which has three values – t, d, and ed. The values of this feature could be computed from the phonological properties of the root and would unify with the same feature in the Regular-Past-Tense-Verb construction. The two solutions presented above, however, are the ones that are most heavily based on morpho-phonology, and as such are of greatest interest to the purpose of the current paper.

**4. Different means of representation**

Aside from the meaningfulness of affixes, the preceding discussion of the English past tense also raises the issue of the precise relation between morphologically related words. Recall from the introduction that various different mechanisms, including inheritance, constituency, and mapping are available in ECG for morphological representation. We will see how each of these works through a case study of the English past tense.

Let us disregard for this section the question of allomorphy in English past-tense suffixes, since our current interest is in demonstrating the use of different representational mechanisms for morphologically related forms, rather than a complete analysis of English past tense verbal morphology. As is clear from the above discussion, there are many possible approaches to representing this paradigm, but we will make the following assumptions, within which we can describe a relatively restricted set of analyses.

1. There is a general REGULAR-PAST-TENSE-VERB construction, which includes in it a form element that represents the suffix. It will therefore maximally have one constituent - a root. (So the general class of solutions we will be considering follow the one in Figure 15, rather than Figure 10.)
2. There are numerous constructions that are available as verbal roots. These are morphological constructions that are specific to particular lexical items, such as *toss*, and pair the form of the root with its meaning. They all inherit from a general VERB-ROOT construction.
3. There are two general verbal constructions of interest to us - the REGULAR-PAST-TENSE-VERB construction and the REGULAR-PRESENT-TENSE-VERB construction.

### 4.1. Four types of relation

The exact relationships among these constructions depends on the mechanism hypothesized to account for their relations. We can give the four interpretations discussed in the introduction to the relation between the VERB-ROOT and the tensed verb constructions. These mechanisms may be used by themselves or in combination.

1. *Stored representation*: the root and other forms overlap in some ways but are structurally unrelated at the constructional level – this is expressed by simply enumerating the various inflected forms as sisters to each other and the VERB-ROOT construction, and may apply to suppletive relations or stored representations. For example, the various verb forms, such as the present form and the past form might inherit from the same parent but bear no direct relations to each other or the root, as seen in Figure 19.

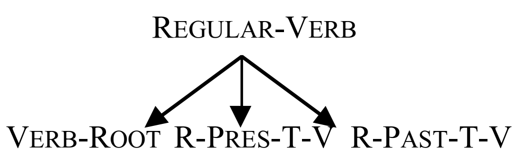


Figure 19 - No direct relation between the inflected forms and the root

2. *Inheritance*: the VERB-ROOT may alternatively inherited by the specific inflected forms that are based on it. As seen in Figure 20, a root in this version inherits from REGULAR-VERB and is inherited by the inflected forms that are based on it.

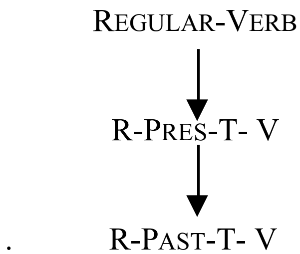


Figure 20 - Inflected forms inherit their root

3. *Constituency*: the root is contained as a constituent of the inflected forms that are based on them. Inheritance and constituency may be notational variants of one another, which differ in this case only in terms of whether an inflected form is claimed to be a more specific version of the root or rather contains it. Figure 21 shows how a root is a constructional constituent of the inflected forms it is part of.

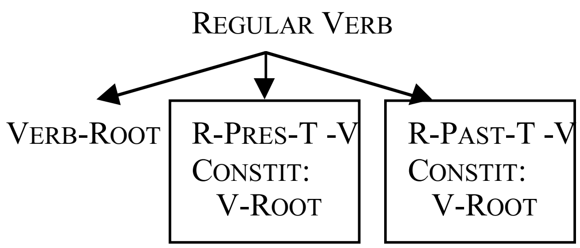


Figure 21 - Roots are constituents of inflected forms

4. *Maps*: the inflected forms are related to one another through maps In Figure 24, a map relates the REGULAR-PRESENT-TENSE- VERB to the REGULAR-PAST-TENSE-VERB.

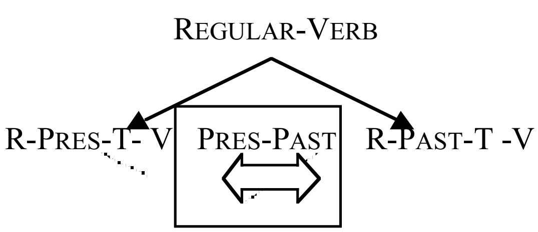


Figure 22 - Roots and inflected forms are related through maps

**4.2. A closer look**

Now that we've seen the general types of relation that can be posited between morphologically related forms, we can look at how they are expressed formally. We show four versions of the PAST-TENSE-INFLECTED-VERB construction below, corresponding to four possibilities outlined in the previous sub-section.

The *stored representation* version involves the inflected constructions inheriting the general characteristics of regular verbs from the REGULAR-VERB construction. The inflected forms then each of independently fill in its details, as the REGULAR-PAST-TENSE-VERB construction does in Figure 23, below.

```
construction REGULAR-PAST-TENSE-VERB
    subcase of REGULAR-VERB
    constructional
        constituents
            root : VERB-ROOT
    form
        elements
            suffix ← /d/
        constraints
            root_f meets suffix
    meaning
        pred : Predication
        pred ↔ root.pred
        pred.aspect ← encapsulated
        pred.setting.time ← past
```

Figure 23 - Regular-Past-Tense-Verb construction, with a single constituent

In this analysis, the Regular-Present-Tense-Verb construction is related to the Regular-Past-Tense-Verb construction (shown in Figure 24) only in that they share a parent, as well as a constructional constituent and some components of their meaning. But there is no direct constructional relation between them - they could be entirely different.

```
construction REGULAR-PRESENT-TENSE-VERB
    subcase of REGULAR-VERB
    constructional
        constituents
            root : VERB-ROOT
    meaning
        pred : Predication
        pred ↔ root.pred
        pred.aspect ← ongoing
        pred.setting.time ← present
```

Figure 24 - Regular-Past-Tense-Verb construction, with a single constituent

Notice that in their relations to the VERB-ROOT construction, these constructions are examples of *constituency* between morphologically related constructions. Each of the inflected constructions contains an instance of the VERB-ROOT construction as a constructional constituent. In general, constituency is used for those cases where a construction in its entirety plays a role in a larger construction.

By contrast, when two constructions are related through *inheritance*, the inheriting construction is seen as a more specific version of the inherited construction. For example, in Figure 25, the REGULAR-PAST-TENSE-VERB construction inherits the REGULAR-PRESENT-TENSE-VERB construction (which can be seen in Figure 24). From this inheritance, it gets a constructional constituent, root, whose form it can constrain to immediately precede the suffix form. Notice that inheritance in the ECG formalism is slightly more flexible than classical inheritance, in that subcases can express constraints on elements that they have inherited from their parent, which are incompatible with the constraints in the parent. For example, in the construction in Figure 25, the setting time of the predication is set to past, whereas in the inherited REGULAR-PRESENT-TENSE-VERB construction, it was assigned the value present.

**construction** REGULAR-PAST-TENSE-VERB
  **subcase of** REGULAR-PRESENT-TENSE-VERB
  **form**
    **elements**
      suffix ← /d/
    **constraints**
      $root_f$ *meets* suffix
  **meaning**
    pred.aspect ← encapsulated
    pred.setting.time ← past

Figure 25 - REGULAR-PAST-TENSE-VERB, inheriting the REGULAR-PRESENT-TENSE-VERB

In the final type of mechanisms for relating morphologically related words, *mapping*, there is a structured map that relates properties of one construction to properties of the other. If we take as a starting point the unrelated constructions in Figures 23 and 24, we can see how the map shown in Figure 26 states the relation between them.

**map** PRESENT-PAST-TENSE-VERB
  **roles**
    present : REGULAR-PRESENT-TENSE-VERB
    past : REGULAR-PAST-TENSE-VERB
  **pairs**
    $present_c$.root ↦ $past_c$.root
    $present_m$.pred ↦ $past_m$.pred
  **constraints**
    $present_c$.root ↔ $past_c$.root
    $past_f$.suffix ← /d/
    past.$root_f$ *meets* $past_f$.suffix
    $past_m$.pred.setting.time ← past
    $past_m$.pred.aspect ← encapsulated

Figure 26 - Map between regular present and past tense verb constructions

Maps identify roles, which can be schemas, as seen in Chang et al. (2002) or can alternatively be constructions, as in Figure 26. Here, the map will operate over two constructions, present and past, which must be instances of the REGULAR-PRESENT-TENSE-VERB and REGULAR-PAST-TENSE-VERB constructions, respectively. The map says that the two constructions in question are mapped onto each other. The constructional constituent root and the meaning role predicate of the REGULAR-PRESENT-TENSE-VERB construction are mapped onto their counterparts in the REGULAR-PAST-TENSE-VERB construction in the **pairs** block. Under **constraints**, those dimensions along which the past tense construction differs from the present tense construction are enumerated. These include the presence of a form element, the suffix, an ordering constraint on this suffix with respect to the root's form (which it must follow), and its tense and aspect specifications.

## 5. Conclusions

The goal of this manuscript has been to lay out the range of mechanisms for morphological analysis within the ECG framework. We have picked certain analyses, but except where noted, these are not meant to be taken as claims about the nature of the internal representation by particular individuals of linguistic knowledge. Rather, they are instructive hypothetical analyses, variants of which may turn out to be particularly useful for understanding the human morphological capacity. For example, other dimensions of abstraction exist beyond those mentioned already. In an inheritance hierarchy with multiple inheritance, like our constructicon, a given morphologically specific construction may inherit from a number of parents, including those within its inflectional class, and those that share features with it across inflectional classes. We have only shown a small piece of the puzzle, demonstrating how all regular past tense verb constructions could inherit from an abstract REGULAR-PAST-TENSE-VERB construction. However, certain aspects of their structure, such as the fact that the plural suffix follows the root, may well derive from a more general Stem-Suffix construction. There might also be particular subclasses of regular past tense verbs (such as those that end in alveolar stops (see Section 3), which are represented in between the most abstract and lexically specific constructions.

Throughout, we have advocated maximal expressiveness of the ECG formalism. The aim for a maximally expressive formalism in which to embed linguistic theories runs opposite to the dominant trend in linguistics, which is to constrain formalisms such that theories using those formalisms are maximally predictive. The idea that a constrained formalism leads to a more predictive (because more constrained) theory seems to arise from the confusion of a formalism with a theory expressed in that formalism. Formalisms are not the same thing as theories using those formalisms, and while theories based on more restricted formalisms are often less likely to "overpredict" linguistic phenomena, they may also lead to misrepresentations of linguistic data for the sole purpose of making them fit in the mold of the restricted model (the "if all you have is a hammer" problem), or worse, may lead to redefinitions of the empirical base to exclude any phenomena that the model in questions cannot capture because of its very restrictions (see for example, the exclusion of processing (or "performance") considerations from most morphological or syntactic models. Grossly, then, the idea behind a maximally expressive formalism is that artificial restrictions on a formalism lead to arbitrarily (and over-)constrained formalisms that are not capable of expressing what humans seem to know about the morphological (or other) complexities of their language.

By contrast, even an extremely expressive formalism can be used to define a quite tightly predictive theory. Consider the relation between physics and mathematics as analogous to the relation between a linguistic theory and a formalism in terms of which it is expressed. Physical phenomena are often captured in mathematical terms. For example, a theory of the acceleration of objects under the sole influence of terrestrial gravity might say that objects have a velocity that is a product of the acceleration of gravity (9.8m/s/s) and the amount of time it has been falling. This theory can be expressed in mathematical terms as:

velocity = 9.8m/s/s * seconds

If our mathematical formalism did not allow for multiplication of rational numbers, but only, say, subtraction of natural numbers, then any theory based on it would certainly be more constrained, but would fail to capture the important generalization that the multiplication in the above theory, which is indeed tightly predictive, requires.

It is our belief that the scientific study of morphology, and of grammatical knowledge in general, should follow the same course as any other scientific enterprise. (1) Observation and description of human linguistic behavior is followed by (2) construction of a hypothesized explanation for the behavior, an explanation which (3) makes predictions about future behavior or observations, which are subsequently (4) tested. A maximally expressive formalism maximally allows the linguist, during the observation phase, to describe whatever it is that humans seem to be doing. If the formalism is sufficiently explicit for theories to be expressed in terms of it, then it also allows that linguist, during the hypothesizing stage, to construct a model of the human as they were performing the behavior in question, such that further predictions can be generated and tested. Expressive formalisms like the one described in this paper thus contribute to the advancement of the scientific study of human linguistic behavior.

## References


Bergen, B. and N. Chang. In Press. Embodied Construction Grammar in Simulation-Based Language Understanding. In J.-O. Östman and M. Fried (eds.), *Construction Grammar(s): Cognitive and Cross-Language Dimensions*. Johns Benjamins.

Bergen, B. and N. Chang (2000). Spatial Schematicity of Prepositions in Neural Grammar. A paper presented at The Fifth International Conference on Conceptual Structure, Discourse and Language, CSDL-V, May 11-14, University of California at Santa Barbara.

Chang N., J. Feldman, R. Porzel and K. Sanders. (2002). Scaling Cognitive Linguistics: Formalisms for Language Understanding. Presented at Scanalu 2002.

Chang, N. and I. Fischer (2000). Understanding Idioms, presented at KONVENS 2000 / SPRACHKOMMUNIKATION in Ilmenau, Germany.

Chang N., S. Narayanan, and M. Petruck (2002). From Frames to Inference. Presented at Scanalu 2002.

Dodge E. and A. Wright. 2002. 'Herds of wildebeests, flasks of vodka, heaps of trouble: An embodied constructional approach to English measure phrases ' In Proceedings of BLS.

Maia, T and N. Chang (2001). Grounding the Acquisition of Grammar in Sensorimotor Representations. Presented at the 2001 AAAI Spring Symposium on Learning Grounded Representations. Stanford, CA.

Wechsler, Stephen and Larisa Zlatic: 2000. A Theory of Agreement and its Application to Serbo-Croatian. Language 76: 799-832.