Overtensing within Optimality Theory

Joseph Paul Stemberger
Department of Linguistics
University of British Columbia

November 20, 2001

1. Introduction

In this paper, I address errors that have become known in the developmental literature as OVERTENSING errors, such as:

I didn't **broke** it. 'I didn't break it.'

Did it **fell**? 'Did it fall?'

In prototypical examples of overtensing errors, a verb that appears with the past-tense auxiliary *did* should appear in the base form but erroneously appears in the past tense. (In non-prototypical, and less-common, errors, the verb occurs with another auxiliary such as *could*, or embedded under a tensed main verb, such as *heard* or *decided*.) Adults also make such errors in spontaneous speech (Stemberger, 1982), though at much lower rates than children do. This paper explores how to derive the characteristics of overtensing errors within Optimality Theory (OT). It is not the purpose of this paper to present the evidence that demonstrates these characteristics. That evidence comes from two studies: (a) experimental studies with adults reported in Stemberger (under review) and Stemberger and Setchell (in preparation), and (b) a recently completed study of spontaneous speech in normally-developing children using the ChiLDES database in Stemberger (in preparation). (Preliminary results from a study of children with Language Impairment also show these characteristics). Rather, this paper explores theoretical mechanisms in OT for explaining the observed patterns.

Overtensing errors have a number of interesting characteristics. (1) They are sensitive to the difference between regular and irregular verbs, and generally occur more often with irregular verbs. (2) They are sensitive to phonological factors, and generally occur more often with verbs that have phonological characteristics that are preferred in English. This is true for both regular and irregular verbs. (3) They are sensitive to lexical frequency, and generally occur more often on low-frequency verbs. These characteristics hold for both child language (where the overall error rate on irregular verbs may approach 25% but is usually much lower) and for adult language (where error rates in experiments can be pushed as high as 6% for the most susceptible verbs). It should be noted that overtensing errors have most often been addressed as purely syntactic errors (e.g. Hurford, 1975; Kuczaj, 1976b; Fay, 1978; Maratsos & Kuczaj, 1978; Pinker, 1984; Andrews, 1990; Roeper & Hollebrandse, 1995),

though Pinker (1984) notes that this does not account for the effect of (ir)regularity. The presence of lexical frequency effects increases the challenge to a syntax-only approach. The presence of phonological effects suggests that syntax is only part of the story.

I address how to derive these characteristics, using Optimality Theory (OT). The goal of OT is to account for effects within a given language, but to situate that explanation within a framework that can account for the range of variation across languages. I begin with observations about tense-marking and introduce morphosyntactic constraints, then explore effects of (ir)regularity, phonology, and frequency. I conclude with some comments about phonology-syntax interactions.

2. Constraints on tense marking

Observation 1:
In English, past tense is expressed on the main verb: **walked**, **broke**, **sang**, etc.

constraint: **Expressed$_{Verb}$(Past)**

Bresnan (2001) uses the cover term **Faith** to govern the expression of morphosyntactic features, without spelling out what sort of faithfulness is involved. One might be tempted to use **MAX(Past)**, which requires the feature {Past} to be in the output if it is in the input. However, for reasons that will become clear below, I am assuming that the output is phonological in nature, and it is not clear that morphosyntactic features should be a part of the phonological representation. Bernhardt and Stemberger (1998) propose the constraint **Expressed(Past)**, which requires that there be something in the phonology that results when {Past} is in the input.

Observation 2:
In English, past tense is expressed on auxiliaries: **did, was, could**, etc.

constraint: **Expressed$_{Aux}$(Past)**

Observation 3:
In English, when both an auxiliary and a main verb are present in a clause, past tense is marked *only* on the auxiliary. This occurs with the auxiliary **did**, in all uses: negative, interrogative, and emphatic:

He didn't sing.	*He didn't **sang**.
Did he sing?	*Did he **sang**?
He DID sing!	*He DID **sang**.

Note that equivalent sentences in closely related languages such as Swedish do not use an auxiliary, and tense is marked on the main verb:

Han **sjöng**.	'He sang.'
Han **sjöng** inte.	'He did sing.'
Sjöng han?	'Did he sing?'
Han **SJÖNG**!	'He DID sing.'

In Swedish, main verbs in negative, interrogative, and emphatic uses can be marked with past tense. This holds true even in English in constructions in which the auxiliary *did* is not used:

He never sang again.	(negative)
He sang OK?	(interrogative)
He SANG!	(emphatic)

A constraint is needed to prevent such marking when an auxiliary is present.

This can be viewed as an instance of blocking, where marking on the auxiliary blocks marking on the main verb; see Andrews (1990). Bock (1982) argues that blocking is part of a more general property of speech: a given piece of semantic or pragmatic information tends to be expressed lexically only once. In this instance, the semantic feature {Past} tends to be expressed in only one place. Some variants of OT use a constraint called **NoMultipleCorrespondence** (which cannot be used here, if morphosyntactic features are not present in the phonological output). Bernhardt and Stemberger (1998) use the following constraint:

constraint: **SinglyExpressed(Past)**

When unviolated, this constraint ensures that past tense will not encoded on both the auxiliary and the main verb.

These three constraints lead to a typology of three different language types, depending on which of the three constraints is ranked lowest. Please note that the relative ranking of the two higher constraints is not relevant, and that the opposite order leads to the same results.

Type 1: The lowest-ranked constraint is the one that limits expression of {Past} to a single token. As a result, past tense is marked on the both the verb and the auxiliary. (I follow the constraint table format of Bernhardt & Stemberger, 1998. In order to avoid radical abbreviation of constraint names, constraints head rows rather than columns. Competing output candidates head rows. Special borders are placed around winning candidates and around cells containing fatal violations.)

sing, Past, Neg	didn't sing	didn't sang	don't sang
Expressed$_{Verb}$(Past)	*!		
Expressed$_{Aux}$(Past)			*!
SinglyExpressed(Past)		*	

Note that, although this ranking is incorrect for adult English for past tense, it is arguably the correct ranking for {Perfect}, which is expressed on both the auxiliary and the main verb: **has walked, has sung, has gone, etc.**

Type 2: The lowest-ranked constraint is the one that requires expression of {Past} on the main verb. The ranking of other constraints allows only a single token expressing {Past}, and requires that {Past} be expressed on the auxiliary. As a result, {Past} is expressed on the auxiliary but not on the main verb.

This is the ranking required for (adult, correct) English.

sing, Past, Neg	didn't sing	didn't sang	don't sang
Expressed$_{Aux}$(Past)			*!
SinglyExpressed(Past)		*!	
Expressed$_{Verb}$(Past)	*		

Type 3: The lowest-ranked constraint is the one that requires expression of {Past} on the auxiliary. The ranking of other constraints allows only a single token expressing {Past}, and requires that {Past} be expressed on the main verb. As a result, {Past} is expressed on the main verb but not on the auxiliary.

sing, Past, Neg	didn't sing	didn't sang	don't sang
Expressed$_{Verb}$(Past)	*!		
SinglyExpressed(Past)		*!	
Expressed$_{Aux}$(Past)			*!

If this option never occurs across languages, OT allows for universal restrictions on ranking, such that **Expressed** would be universally ranked higher for auxiliaries than for main verbs. However, occasional tokens of this sort of construction are observed in child language.

Young children often produce overtensing errors, in which {Past} is expressed on both the auxiliary and the main verb (e.g. Hurford, 1975; Kuczaj, 1976b; Fay, 1978; Maratsos & Kuczaj, 1978; Pinker, 1984; Roeper & Hollebrandse, 1995), though the error rate is rarely more than 20% during the period of maximal use. Adults also occasionally make such errors, at a very low rate in natural speech, which can be pushed as high as 6% of trials in experimental situations for the most susceptible verbs. Such errors can be derived if the rankings of English are unstable. Although **SinglyExpressed(Past)** is generally ranked higher than **Expressed$_{Verb}$(Past)**, sometimes the ranking is in the opposite order. (See Bernhardt & Stemberger, 1998, for the approach to variability taken here. See Boersma, 1998, for a similar approach.)

Although this ranking yields overtensing errors, it does so in a way that predicts that overtensing errors will occur at the same rate for all verbs. This is not the case. (a) Overtensing is most likely to occur with irregular verbs. Errors do occur on regular verbs, but at a lower rate. (b) There are lexical frequency effects. Low frequency verbs are much more likely to be involved with overtensing than are high frequency verbs. (c) There are phonological effects, of three sorts. (i) If the base and past tense forms have different vowels, inherent phonological differences between those vowels lead to biases towards one vowel or the other. (Stemberger, 1993, termed this VOWEL DOMINANCE in the context of overregularization errors in child language.) If the dominant vowel is in the past tense forms, overtensing is more likely. (ii) Stemberger and Treiman (1986) and Stemberger (1990) note that, when a consonant competes with nothing, the consonant tends to win (and thereby is produced). When a past tense form has a consonant corresponding to nothing in the base form, there is a tendency for the past tense form to be produced in overtensing errors. (However, this appears to be a much smaller effect than vowel dominance.) (iii) Errors involving regular verbs occur at a higher rate if the resulting rime fits the constraints on rimes that are allowed in monomorphemic words in English.

It is necessary to expand this basic account in order to deal with differences between different types of lexical items.

3. The Regularity Effect

There must be more than one constraint dealing with the expression of past tense in English, since there are multiple patterns. One pattern (*-ed*) is the default, and is used with large numbers of verbs, including most newly created verbs. All other patterns are referred to as IRREGULAR. There are many different ways to deal with these irregular patterns, and the details need not concern us here. I will

make use of a short-hand device, listing the entire irregular form in the constraint. (See also Russell, 1999.) Such constraints are to be interpreted as e.g.: "{Past} in combination with /sɪŋ/ must be realized as [sæŋ]." In order for irregular patterns to be output successfully, these constraints must be ranked higher than the constraint for the default pattern. This is exactly parallel to the ranking of faithfulness constraints for default and nondefault phonological features, where nondefault (**Survived(Dorsal)** or **Ident(Dorsal)** or **MAX(Dorsal)**, depending on the variant of OT used) must be ranked higher than default (**Survived(Coronal)** or **Ident(Coronal)** of **MAX(Coronal)**).

Past(sæŋ) » Past(d,Right)

Note the effects of variability of ranking. If **Past(d)** is ranked higher than **Past(sæŋ)** on any trial, then the output must have the default suffix *-ed* added, without the vowel change. An overregularization error results: **singed, choosed**, etc.

It is easy to derive frequency effects on overregularization errors. Bernhardt and Stemberger (1998) propose that faithfulness constraints are ranked higher for high-frequency forms than for low-frequency forms. Thus, the relative ranking of the irregular constraints should correlate with frequency: higher for higher-frequency past tense forms. The higher the frequency, the greater the separation in ranking from the default constraint, the less likely it is that the rankings will ever be in the opposite order, and so the lower the error rate.

Note also that, in order to get the /æ/ of *sang*, the output must be unfaithful to the /ɪ/ of the base *sing*. As a result, **Past(sæŋ)** must be ranked higher than faithfulness constraints relevant to /ɪ/: **Survived(+high)** and **Survived(-low)**, for this verb. With higher ranking of the phonological faithfulness constraints of the base verb, the vowel [ɪ] would result in the output, thereby preventing the irregular past tense form *sang* from appearing.

The regularity effect on overtensing errors can now be derived. First, consider that overtensing errors are in the minority, even for children during their maximal period of usage. This implies that **SinglyExpressed(Past)** is ranked fairly high, but that it is sometimes ranked low enough so that it is below other constraints. Second, recall that **Past(*IRREGULAR*)** constraints must be ranked higher than the default **Past(d)** constraint. This leads us to conclude that, most of the time, the following ranking holds:

	sing, Past, Neg		walk, Past, Neg	
	didn't sing	didn't sang	didn't walk	didn't walked
SinglyExpressed(Past)		*!		*!
Past(*IRREGULAR*)	*			
Past(d)			*	

Sometimes, **SinglyExpressed(Past)** is ranked lower. The ranking stays close to the mean ranking, however. It is more likely to be ranked closer to the mean than farther away. As a result, the following ranking is next most likely:

	sing, Past, Neg		walk, Past, Neg	
	didn't sing	didn't sang	didn't walk	didn't walked
Past(IRREGULAR)	*!			
SinglyExpressed(Past)		*		*!
Past(d)			*	

Note that the high ranking of the irregular constraints leads to overtensing in this instance, while the same ranking prevents overtensing for regular verbs. It is less likely that **SinglyExpressed(Past)** will be ranked even lower. Note, however, that when it is, both regular *and* irregular verbs take part in overtensing:

	sing, Past, Neg		walk, Past, Neg	
	didn't sing	didn't sang	didn't walk	didn't walked
Past(IRREGULAR)	*!			
Past(d)			*!	
SinglyExpressed(Past)		*		*

Since all rankings that result in the overtensing of regular verbs also result in the overtensing of irregular verbs, but some rankings that result in the overtensing of irregular verbs do not allow overtensing of regular verbs, this analysis predicts that irregular verbs are more often overtensed than regular verbs. This is the case.

4. Phonological Effects I: Vowel Dominance Effects

As noted above, there are effects due to competition between the base vowel and the vowel of the past tense. In order to output the vowel of the past tense form, the output must be unfaithful to the vowel of the base form. This is only possible if the constraint that forces the use of the irregular past

tense form is ranked higher than the relevant faithfulness constraint to the vowel. If this is not the case, then the vowel of the base results. In the case of overregularization errors, if the base vowel is in the output, the default past tense pattern then comes into use (Stemberger, 1993): e.g. **singed, breaked, falled**. In the case of overtensing, a correct form generally results, with past tense marked on the auxiliary only.

In pre-OT work, Stemberger (1993) noted that there is a general phonological bias towards nondefault features. In OT terms, faithfulness is higher to nondefault features than to default features. In (adult) English, the following defaults seem to hold:

default	nondefault
-high	+high
-low	+low
-back	+back
Coronal	Labial
	Dorsal
μ	μμ

The default vowel for English is /ɛ/ (see Stemberger, 1992a, for arguments). Note also that, by the Addition Bias noted above, a monophthong such as /iː/ is unmarked compared to a diphthong such as /ɑɪ/. When the vowel of the past tense form has only nondefault features (when different) compared to the base vowel, the prediction is that the vowel of the past tense form will appear in the output (when the expression of past tense is appropriate). However, when the past tense form has only nondefault features (when different), the prediction is that the vowel of the base will appear in the output, and that an overregularization error will result. This leads to many regularization errors like **falled** (past tense has default [-low,-back]) but few like **getted** (past tense has nondefault [+low,+back]). This is the strongest predictor of overregularization rates in the speech of young children (Stemberger, 1993; Marchman, 1997).

For overtensing errors, the same effect should be present. However, the results should be the opposite. Past tense forms with nondefault features should be *more* prone to overtensing, because the nondefault features are in the output. In the ranking in the following constraint table, overtensing will not normally occur, except when forced by phonological faithfulness effects:

	gɛt, gɑt, Past, Neg		fɑl, fɛl, Past, Neg	
	didn't gɛt	didn't gɑt	didn't fɑl	didn't fɛl
Survived(+back/+low)	*!			*!
SinglyExpressed(Past)		*		*
Expressed$_{Verb}$(Past)	*		*	
Survived(-back/-low)		*	*	

Note that variability of ranking for **SinglyExpressed** is *predicted* to lead to interactions with phonology in this way. In this way, OT is preferable to any theory in which such interactions are not predicted to occur, but in which a new mechanism can be introduced for the sole purpose of dealing with such effects post-hoc.

Note, however, that OT actually predicts a typology of two types of languages. One type is the sort just discussed. A second type is one in which *all* of the phonological faithfulness constraints are ranked *very low* relative to the morphological constraints, such that no phonological faithfulness constraint would ever be ranked above the morphological constraints. (This seems to be the assumption of earlier linguistic theories such as Chomsky and Halle, 1968, in which morphology was ordered before phonology.) In such a language, these vowel effects would not be present. When dealing with child language, we might also expect to find those two sorts of systems. However, the system with vowel dominance effects seems to be more common.

5. Phonological Effects II: The Addition Bias

Some irregular verbs have more consonants in the past tense form than in the base form. In a few verbs, this is the only difference: **burnt** (-*t* added to base). In North American English, *burnt* is the only such past tense form, and it varies with the regular variant *burned*. In most cases, there is additionally a vowel change: **lost** (extra consonant /t/ at the end of the word, with change of the vowel), **went** (two extra consonants /nt/ at the end of the word, with difference in the vowel). This phonological difference between base and past tense form may also have an effect on output.

Stemberger and Treiman (1986) and Stemberger (1990) addressed errors in adult speech in which there was interference between two words, one of which started with a single consonant and one of which started with a consonant cluster. In a phrase such as *bite block*, a common error is one in which the /l/ from the second word is added to the first word: *b**L**ite block*. It is comparatively rare for a consonant to delete: *bite b**Ø**ock*. Stemberger and Treiman argue for a general phonological bias that resolves competition between a consonant and nothing in favor of the consonant. Stemberger (1989)

shows that this bias also holds true for speech errors in the speech of young children. In OT terms, it appears to be faithfulness to the presence of a consonant. Bernhardt and Stemberger (1998) use the following constraint (in place of MAX):

Survived(C-Root): If a consonant is present in the input, it should also be present in the output.

Interactions of the Consonant Addition Bias with vowel dominance and overtensing are shown in the following constraint tables:

	sliːp, slɛpt, Past, Neg		spiːd, spɛd, Past, Neg	
	didn't sliːp	didn't slɛpt	didn't spiːd	didn't spɛd
Survived(C-Root)	*!			
Survived(+high)		*		*!
SinglyExpressed(Past)		*		*
Survived(-high)	*		*	

	breɪk, broʊk, Past, Neg	
	didn't breɪk	didn't broʊk
Survived(C-Root)		
Survived(+back)	*!	
SinglyExpressed(Past)		*
Survived(-back)		*

The first table shows that the presence of the final /t/ in verbs like *slept* (including in the suppletive form *went*) leads to overtensing, even though the same ranking prevents overtensing in simple vowel-changing irregular verbs with the same two vowels. The second table shows that this sort of ranking can still result in overtensing in simple vowel-changing verbs, if faithfulness for the nondefault vowel in the past tense form is ranked high enough. Note, however, that a third ranking can lead to overtensing on verbs like *slept* but not on verbs like *broke*:

	sliːp, slɛpt, Past, Neg		breɪk, broʊk, Past, Neg	
	didn't sliːp	didn't slɛpt	didn't breɪk	didn't broʊk
Survived(C-Root)	*!			
SinglyExpressed(Past)		*		*!
Survived(+back)			*	
Survived(-back)				*

Thus, it is possible that a child might not show an effect of vowel dominance on overtensing, but would show an effect of consonant addition. Other children would show an effect of both factors.

Note that there are two verbs in English that should show a bias towards the base form, according to the consonant addition bias. Both contain a nasal in the base form that is absent from the past tense form: **stood** (cf. base *stand*) and **thought** (cf. base *think*). There is insufficient data to know how these verbs behave relative to overtensing.

6. Phonological Effects III: Consonant feature biases

There are several verbs that, in addition to having past tense vowels that are dominant over the vowel of the base, have consonant changes with the opposite effect.

brought	*bring*	past /t/ vs. base /ŋ/
caught	*catch*	past /t/ vs. base /č/

Stemberger (1991) has shown that, in speech errors, there is a tendency for coronals to be replaced by velars and by palatoalveolars, and that there is a tendency for stops to be replaced by nasals. As with vowels, the phonological bias favors nondefault features ([Dorsal, -anterior, +nasal]) at the expense of defaults ([Coronal, +anterior, -nasal]).

In principle, the consonantal feature biases could create a bias towards the consonants of the base form, preventing overtensing errors from occurring, even though other phonological biases favor the vowel of the past tense form (and thus make the forms prone to overtensing). There is insufficient data on this point. However, in general, these verbs appear to have fairly high overtensing rates, and thus show effects of vowel dominance rather than consonant dominance. This suggests that the vowel faithfulness features are ranked higher than the consonant faithfulness features. (This is compatible with the observation that the full vowel system tends to be mastered earlier in child speech than the full

consonant system. See Bernhardt & Stemberger, 1998, for discussion.)

7. Phonological Effects IV: Effects on regulars

Although children have a lower rate of error on regular verbs, they do produce such errors. In this study, the error rate was about 10%. However, there are differences between different types of verbs, based on their phonological characteristics. (The effect discussed here has never been addressed for overtensing errors in adult speech.)

First, no errors in this study were observed on verbs that end in /t/ or /d/. It has been shown that children often show no-marking errors on such verbs, producing base forms such as **want** and **glide** instead of the past tense forms *wanted* and *glided* (e.g. Berko, 1958; Bybee & Slobin, 1982). If a child fails to add *-ed* in such regular words, the child will also be less likely to overregularize irregular verbs that end in /t/ or /d/ (Maratsos, 2000). The lack of overtensing errors on verbs that end in /t/ and /d/ is presumably a reflection of the same underlying problem. This is simply a third manifestation of the problem.

Second, there is an effect of the nature of the word ending that is created when the [t] or [d] of the *-ed* suffix is added. There are two important categories, for our purposes.

(h) the resulting rime is one that occurs generally in English, including in monomorphemic forms (see Hammond, 1999, for recent discussion)

Vd:	cried	(cf:	*side*)
Vft:	coughed	(cf:	*soft*)
	knocked	(cf:	*concoct*)
VVld:	peeled	(cf:	*field*)

In general, such rimes involve a short vowel followed by two consonants. If the last consonant is /t/ or /d/, a preceding nasal must be of the same place of articulation, as in /nt/ and /nd/, but */mt/ and */md/). Alternatively, they can be a long vowel or diphthong followed by a single consonant, or by two consonants if both are coronals (/st, nt, lt, nd, ld/). These clusters occur in monomorphemic words, and either have a reasonable frequency outside of inflected forms, or closely resemble clusters that are common.

In contrast, there are other past tense forms that create rimes that in principle cannot occur in a monomorphemic word:

VCC: rammed (*/md/)

VVCC: liked (*/aɪkt/)

VCCC: danced

These generally contain a long vowel or diphthong followed by two consonants where one is not a coronal, or a short vowel followed by a non-occurring consonant cluster or by an overly long consonant cluster. These have no frequency outside of inflected forms, and so are not produced until inflections are acquired.

Evidence suggests that children's first regular past tense forms have clusters (and rimes) that are already present in the child's speech in monomorphemic forms (Bernhardt & Stemberger, 1998). Non-basic clusters develop later. In OT, it is not necessarily the case that such inflectional clusters, when they finally appear, are fully on a par with more basic clusters. The theory holds that there can still be subtle ways in which clusters are avoided.

One subtle way to avoid them would be to avoid overtensing errors involving them. This is in fact the case. In the present study, children produced overtensing on regular forms with basic rimes with about an 18% error rate. They produced overtensing errors on regular verbs that resulted in non-basic rimes with only a 4% error rate. This is easily derived in OT:

	lock, Past, Neg		like, Past, Neg	
	didn't lɑk	didn't lɑkt	didn't lɑɪk	didn't lɑɪkt
NotComplex(Nucleus) &NotComplex(Coda)				*!
Expressed$_{Verb}$(Past)	*!		*	
SinglyExpressed(Past)		*		*

The standard constraint in the first row of the constraint table is a complex constraint, that rules out forms that have both a branching nucleus (with a long vowel or diphthong) and a branching coda (with two or more consonants). In general, such rimes are ruled out in English and many other languages (though English allows a few systematic exceptions, especially if both consonants have default coronal place of articulation; see Hammond, 1999, for recent discussion). Ranking this constraint high prevents overtensing errors if the resulting rime would violate the constraint. Other constraints governing e.g. consonant clusters would have the same sort of effect on other verbs.

The constraint ranking shown in this table appears at first to have an undesirable side-effect: it

will erroneously prevent regular past tense marking on *like* everywhere, even when there is no auxiliary. However, the form *liked* does occur for most of these children. It is a common assumption in OT that constraints occur in many variants, with different domains. For **Expressed** constraints, it is reasonable to assume that there is a high-ranked variant **Expressed(Past)** that requires past tense to be marked somewhere. This is supplemented with lower-ranked constraints that determine exactly where past tense will be expressed. If the general **Expressed** constraint is high-ranked, and the main verb is the only place where past tense can be expressed, then it will be expressed on the main verb. The ranking here, then, will have the described effect only on overtensing, where marking past tense on the verb is not required by the high-ranked general **Expressed(Past)** constraint.

This effect shows that regular inflection is subject to phonological constraints. It adds to the acquisition literature on that topic (see Bernhardt & Stemberger, 1998) in that this is not purely a morphological error, but an error with a syntactic component.

Phonological effects on regular forms raise the possibility that *all* observed differences between regular and irregular forms in overtensing might be due to phonological effects. Stemberger (1992b) controlled for phonological effects as closely as possible in experiments with adult speakers of English. He concluded that phonological effects are larger than any direct effect of (ir)regularity, but that there *may* be some effect of (ir)regularity. However, the nature of the irregular verbs in English prevented controlling for the phonology perfectly, and the differences that were observed *may* have been phonological. In this study, error rates on regular verbs with basic rimes were as great as error rates on irregulars that add a consonant (such as *lost*, *said*, and *went*). At this point, it is difficult to argue that there are direct effects of (ir)regularity in addition to effects of phonology. But direct effects of (ir)regularity cannot be ruled out.

8. Frequency Effects

There is a standard frequency effect on overtensing errors, both in adult errors (as shown experimentally) and in the spontaneous speech of children. The system as I have developed it so far predicts that lexical frequency will have an effect. However, it makes the wrong predictions, as shown in the following constraint table:

WRONG RESULTS!!!	break, broke, Past, Neg		wake, woke, Past, Neg	
	didn't break	didn't broke	didn't wake	didn't woke
Past(*IRREGULAR*)$_{HIGH}$	*!			
SinglyExpressed(Past)		*		*!
Past(*IRREGULAR*)$_{LOW}$				*

According to the version of OT used here, certain kinds of constraints are ranked higher for high-frequency lexical items than for low-frequency lexical items. For irregular past tense forms, this means that the irregular form is ranked farther above the default suffix for high-frequency irregular forms than for low-frequency irregular forms. Given variability in ranking, the low-frequency forms are more likely to be involved in errors in which they are ranked below the default suffix *-ed*. As a result, low-frequency verbs are more likely to be overregularized than high-frequency verbs (e.g. Bybee & Slobin, 1982, and Marcus et al., 1992, for child language; Stemberger & MacWhinney, 1986, for adult speech).

However, this same ranking predicts that the constraints governing high-frequency irregular forms are more likely to be ranked above **SinglyExpressed(Past)**, and so are more likely to be involved in overtensing errors. The theory predicts an *anti*-frequency effect (i.e., more errors on high-frequency verbs than on low-frequency verbs), rather than the standard frequency effect. (The same is also likely to be true of the models of past tense production put forward by Rumelhart & McClelland, 1986, and Marcus et al., 1992.) This is a problem, since a standard frequency effect is observed.

In most versions of OT, this seems to be an insurmountable problem. I see no way of deriving this frequency effect, without either finding an entirely different approach that the one taken here (with no certainty that any alternative will work), or deciding that overtensing in principle should not be derived via the grammar.

In the version of OT used here, there is a possible way to derive these frequency effects. Bernhardt and Stemberger (1998) argue that constraint rankings are related to activation levels. This derives observed differences in faithfulness between different positions in the word and syllable. It also predicts that faithfulness is greater in high-frequency words than in low-frequency words. Although high-frequency irregular past tense forms have higher-ranked morphological constraints, the phonological faithfulness constraints for the base vowel are also ranked higher. There is a very strong correlation between high frequency in past tense forms and high frequency in other forms of the verb: relatively high-frequency verbs are relatively high in frequency in all their inflected forms (MacKay, 1976). This is to be expected, if speakers are as likely to speak of particular events regardless of whether past, present, or future time is relevant. Previous work has tended to focus on the high frequency of the *past-tense* form. To derive a standard frequency effect on overtensing, we must focus on the high frequency of the *base* form.

Note that if phonological faithfulness to the base form of high-frequency verbs is too great, it would be the case that the base vowel would be more likely to survive in the output, and overregularization errors would be *more* common on high-frequency irregular verbs than with low-frequency irregular verbs. Since this is not the case, we must assume that, overall, there is little difference between high-frequency and low-frequency verbs in the relative ranking of phonological constraints vs. the morphological constraints that lead to irregular forms. This is plausible, since *both* the morphological constraints *and* the phonological constraints are shifted upwards in the constraint rankings for high-frequency forms.

However, we are dealing here with the interaction of a morphological constraint (**SinglyExpressed**) that is independent of the particular irregular verbs, and that should be unaffected by lexical differences in ranking. All that is needed is a constraint ranking that will allow greater faithfulness to the *base* vowel to prevent overtensing in high-frequency verbs:

	breɪk, broʊk, Past, Neg		weɪk, woʊk, Past, Neg	
	didn't breɪk	didn't broʊk	didn't weɪk	didn't woʊk
Survived$_{Base}$(-back)$_{HIGH}$		*!		
Expressed$_{Verb}$(Past)	*		*!	
SinglyExpressed(Past)		*		*
Survived$_{Base}$(-back)$_{LOW}$				*

In order to get this to work, we must have constraints that differentially apply to the phones of the base form vs. those of affixes. This is a common assumption (Beckman, 1997; Bernhardt & Stemberger, 1998; Alderete et al., 1996). Note that, since base forms are in general more frequent than the irregular past tense form of the same verb (MacKay, 1976), faithfulness is expected to be higher to the base form than to the irregular vowel. (Interactions are actually quite complex. In the case of *broke*, faithfulness to the vowel of the past tense form is faithfulness to the nondefault feature [+back], while faithfulness to the vowel of the base form is faithfulness to the default feature [-back]. Faithfulness is ranked higher to nondefault features than to default features. The relative ranking of **Survived$_{Base}$(-back)** and **Survived$_{Affix}$(+back)** is not *a priori* determinable by the theory. For the explanation here to work, faithfulness to the default feature of the base must be greater than faithfulness to the nondefault feature of the irregular form, at least some of the time.)

By focussing on faithfulness to the base vowel, we derive a standard frequency effect. However, a possible drawback to this solution is that it is just one pattern that OT will allow. If faithfulness to the base vowel is ranked lower, then the theory could derive an anti-frequency effect rather than a frequency effect. The theory seems to predict that some speakers will show a frequency effect, while others will show an anti-frequency effect. There is no evidence for such a difference between speakers. Overall, a frequency effect is detectable in the data for both children and adults, so frequency effects must be more common than anti-frequency effects, at very least.

To eliminate the possibility of speakers showing an anti-frequency effect here, there are two possible approaches. (1) We can hope that it will be predicted by some independent characteristic of learning, the relevance of which to overtensing has yet to be elucidated. (2) We can develop some explanation that will take care of it. One possibility is to question the nature of the input. I have been assuming that {Past} is present in the input to the phrase *didn't break* to the same degree that it is in

the input to *broke*. One could argue, however, that {Past} is present to a lesser degree in *didn't break*, possibly as an affect of competition with *did*. If the activation level of {Past} is less when *did* is present, then all constraints governing expression of past tense are shifted downwards in the ranking, including those governing the presence of the irregular form. If so, interactions with the (unaffected) rankings of faithfulness for base vowels might have the desired effect of guaranteeing that anti-frequency effects never occur. It remains to work out the details.

9. Syntactic effects: questions, negatives, and emphatics

My interest here is primarily with phonological and lexical effects. I will not address syntactic effects in this version of the paper.

One issue that I do wish to address is whether overregularized forms are as common in overtensing as they are in the production of simple past-tense forms. In order to answer such a question, data on both overtensing and overregularization must be available for the same child during a defined period of development, taking into account differences between verbs in the tendency to undergo overregularization and overtensing. I have not been doing such a study, and know of none in the literature.

Estimating chance occurrence of overregularizations in overtensing errors is complex, and must take into account phonological effects on overtensing in regulars and the child's propensity to overregularize a particular verb. It is also hampered by the fact that versions of past tense with the *did* auxiliary occur in child samples at a rate of only 5% to 15% as often as simple past tense forms, so data are more limited. As an example, consider the data from Abe (Kuczaj, 1976a). Abe targeted base-dominant or past-dominant irregulars for negative or interrogative past only 42 times during the period I examined, compared to about 300 simple past tense forms for the same verbs during the same period. Abe overtensed regular verbs with basic rimes at a rate of 11.1%, but did not overtense any other regulars. He targeted base-dominant irregulars whose overregularized form would have basic clusters only 8 times; at the time, he overregularized base-dominant verbs at a rate of 79.6%. We predict that overregularized overtensed forms would occur on 8*.111*.796 trials: 0.71 tokens. For past-dominant irregulars (overregularized at a rate of 14.1% at that age), we predict 14*.111*.141: 0.22 tokens. We predict that Abe would make 0.93 overregularized overtensed errors. He made none, which is not significantly different from chance. Clearly, a much larger study is needed in order to address this issue at all.

10. But is it grammar?

One reaction to the analysis presented here is that perhaps a theory of grammar such as OT should not be expected to account for the observed effects on overtensing. The fact that overtensing is

found on a low percentage of trials (under 20% for most children even during the period with the greatest error rate) could be used to argue that these are performance errors (e.g. Pinker, 1984). Overtensing is much less common in adult speech, and adults clearly view overtensed forms as performance errors. It is not clear that this matters for the analysis of overtensing.

It should be noted that similar comments have been made about other aspects of child language and of adult errors. Overregularization errors (such as *breaked* 'broke' and *choosed* 'chose') are common in child language. It has often been noted that they are subject to frequency effects (e.g. Bybee & Slobin, 1982; Marcus et al., 1992). Marcus et al. also report that overregularizations account for a small percentage of tokens (under 5%) for most children. Stemberger (1993) reports that they are subject to vowel dominance effects. Marcus et al. conclude that overregularizations are a type of performance error.

Note that overregularizations and overtensing seem to be very different from many types of performance error. They relate to the standard grammar of English in a way that errors like the following do not:

Phonological:	Why was **s**ee **sh**urprised?	"Why was ***sh***e ***s***urprised?"
Syntactic:	Unless you have **a magnifying glass**.	"... have ***binoculars***."

In the first sentence, the speaker reversed the word-initial phonemes in two different words; this is unlike any phonological pattern in the grammar of English. In the second sentence, the speaker accessed the wrong lexical item (but one that was semantically related to the target word). These seem to be fully outside of the normal functioning of the grammar. Regularization errors and overtensing errors seem to be quite different: they are slight extensions of the grammar. It can be argued that they are mis-applications of the same grammatical patterns that produce correct sentences.

What is the relation of competence to performance? The view that is often assumed seems to be that performance is unrelated to competence. This implies that the procedures used by speakers and listeners during actual performance bear no resemblance to the descriptions of language that are found in competence grammars. One can ask the question (as have many psychologists) why researchers doing work on language acquisition or adult language processing should pay any attention to competence, if it is unrelated to what speakers and listeners actually do. A more interesting view is that performance is competence plus something else. Performance uses the competence grammar, but shows additional properties that are due to actual processing in real time. This seems to be the viewpoint of all psycholinguists and language acquisition researchers who make use of linguistic theories.

In this view, overregularization and overtensing make use of general grammatical procedures. Although something has gone awry, the grammar nonetheless molds the outcome. Davis (1987) argues that such phenomena (including overtensing) are performance-mediated errors in which the grammar

plays a central role. He maintains that we need to supply a competence grammar that can account for such phenomena. In the OT analysis given here, the rankings that yield overtensing are extremely unlikely, and adults do not like the results. However, they can be viewed as resulting at the extreme endpoints of the range of variability of the rankings of the relevant constraints. This is different in kind from the speech errors given above, which result from extra-grammatical malfunctioning of the lexical accessing or phoneme sequencing procedures of speech production. An analysis of how overtensing arises in OT is quite appropriate, even if overtensing is a variety of performance error.

11. Conclusions

Overtensing errors turn out to be a complex interaction of morphosyntactic (**SinglyExpressed**), lexical (irregularity, frequency) and phonological (faithfulness to nondefaults, to base vowels, and to consonants) effects. I have provided a way to derive these effects within Optimality Theory. The rankings of different phonological faithfulness constraints and of morphological constraints is known independently. In the version of OT assumed here, the effects of frequency on rankings have been proposed to account for independent effects. By variably ranking **SinglyExpressed** below some of these other constraints but above others, all of the observed effects can be derived.

One conclusion that can be drawn is that syntax is not as independent of phonology as was once believed. In previous theories, syntactic and morphological processing were completed before the speaker encoded the words phonologically. Syntax and morphology could have an affect on phonology, but not vice versa. In OT, there are syntactic constraints, morphological constraints, and phonological constraints. The theory gives no reason why *all* syntactic constraints must be ranked higher than *all* phonological constraints (though, of course, this could be arbitrarily stipulated). There are phenomena showing that phonology affects syntax, both in grammars (e.g., Rice & Svenonius, 1998; Broadwell, 2000) and in language acquisition (Stemberger & Bernhardt, 1997). The characteristics of overtensing errors lead to the same conclusions.

References

Alderete, J. , Beckman, J. , Benue, L. , Gnanadesikan, A. , McCarthy, J. , & Urbanczyk, S. (1996). Reduplication and segmental unmarkedness. Rutgers Optimality Archive, ROA-134.

Andrews, A.D. (1990). Unification and morphological blocking. *Natural Language & Linguistic Theory*, **8**, 507-557.

Beckman, J. N. (1997). *Positional Faithfulness*. Doctoral dissertation, University of Massachusetts, Amherst. (ROA-234)

Berko, J. (1958). The child's learning of English morphology. *Word*, **14**, 150-177.

Bernhardt, B.H., & Stemberger, J.P. (1998). *Handbook of phonological development: From the perspective of constraint-based nonlinear phonology*. San Diego, CA: Academic Press.

Bock, J.K. (1982). Towards a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, **89**, 1-47.

Boersma, P. (1998). *Functional phonology.* (Ph.D. dissertation, University of Amsterdam.) The Hague: Holland Academic Graphics.

Bresnan, J. (2001). The emergence of the unmarked pronoun. In G. Legendre, S. Vikner, & J. Grimshaw (Eds.), *Optimality-theoretic syntax*. Boston, MA: MIT Press.

Broadwell, G.A. (2000). On the phonological conditioning of clitic placement in Zapotec. ROA-422.

Bybee, J.L., & Slobin, D.I. (1982). Rules and schemas in the development and use of the English past tense. *Language*, **58**, 265-289.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Cambridge, Massachusetts: MIT Press.

Davis, H. (1987). *The acquisition of the English auxiliary system and its relation to linguistic theory*. Unpublished doctoral dissertation, University of British Columbia.

Fay, D. (1978). Transformations as mental operations: a reply to Kuczaj. *Journal of Child Language*, **5**, 143-149.

Hammond, M. (1999). *The Phonology of English: A Prosodic Optimality-Theoretic Approach*. Oxford: Oxford University Press.

Hurford, J.R. (1975). A child and the English question formation rule. *Journal of Child Language*, **2**, 299-301.

Kuczaj, S. (1976a). *-ing, -s and -ed: A study of the acquisition of certain verb inflections.* Unpublished doctoral dissertation, University of Minnesota.

Kuczaj, S.A. (1976b). Arguments against Hurford's 'Aux Copying Rule'. *Journal of Child Language*, **3**, 423-427.

MacKay, D.G. (1976). On the retrieval and lexical structure of verbs. *Journal of Verbal Learning and Verbal Behavior*, **15**, 169-182.

Maratsos, M. (2000). More overregularizations after all : new data and discussion on Marcus, Pinker, Ullman, Hollander, Rosen & Xu. *Journal of Child Language*, **27**, 183-212.

Maratsos, M., & Kuczaj, S.A. (1978). Against the transformational account: A simpler analysis of auxiliary overmarkings. *Journal of Child Language*, **5**, 337-345.

Marchman, V. (1997). Children's productivity in the English past tense: The role of frequency, phonology, and neighborhood structure. *Cognitive Science*, **21**, 283-304.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., and Xu, F. (1992). *Overregularization in Language Acquisition*. (Monographs of the Society for Research in Child Development). **57** (4, Serial No. 228).

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Rice, C., & Svenonius, P. (1998). Prosodic V2 in Northern Norwegian. Paper presented at the 17th West Coast Conference on Formal Linguistics, Vancouver, BC, February 1998.

Roeper, T., & Hollebrandse, B. (1995). *GALA*.

Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In D. Rumelhart

& J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1 (pp. 216-271). Cambridge, MA: Bransford Books.

Stemberger, J.P. (1982). *The lexicon in a model of language production*. Doctoral Dissertation, UCSD. New York: Garland Publishing Inc. 1985.

Stemberger, J.P. (1989). Speech errors in early child language production. *Journal of Memory and Language*, **28**, 164-188.

Stemberger, J.P. (1990). Wordshape errors in language production. *Cognition*, **35**, 123-157.

Stemberger, J.P. (1991). Apparent anti-frequency effects in language production: The Addition Bias and phonological underspecification. *Journal of Memory and Language*, **30**, 161-185.

Stemberger, J.P. (1992a). Vocalic underspecification in English language production. *Language*, **68**, 492-524.

Stemberger, J.P. (1993). Vowel dominance in overregularization. *Journal of Child Language*, **20**, 503-521.

Stemberger, J.P. (under review). Overtensing and the effect of regularity. Unpublished manuscript, University of British Columbia.

Stemberger, J.P. (in preparation). Phonological effects in overtensing. Unpublished manuscript, University of British Columbia.

Stemberger, J.P., & Bernhardt, B.H. (1997). Syntactic development limited by phonological development. Paper presented at the 38th meeting of the Psychonomic Society, Philadelphia, November 1997.

Stemberger, J.P., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory and Cognition*, **14**, 17-26.

Stemberger, J.P., & Setchell, C.M. (in preparation). Vowel dominance and morphological processing. Unpublished manuscript, University of British Columbia.

Stemberger, J.P., & Treiman, R. (1986). The internal structure of word-initial consonant clusters. *Journal of Memory and Language*, **25**, 163-180.

Vihman, M.M. (1982). The acquisition of morphology by a bilingual child: A whole-word approach. *Applied Psycholinguistics*, **3**, 141-160.