# Frequency Effects on Determining English Secondary Stress

Young-Ah Do
**(Seoul National University)**

**Do, Young-Ah. 2007. Frequency Effects on Determining English Secondary Stress.** *SNU Working Papers in English Linguistics and Language* 6, 54-67. This paper reviews and builds upon previous studies about how word frequency is relevant in determining English secondary stress in derived forms, using corpus-based research. The result of statistical research will incorporates into Optimality Theoretic Grammar. I propose the group-specific approach; if the frequency of the derived forms is high, the stress pattern of the words will be marked at the lexicon independently. If the frequency is low, however, the word tends to depend on base form to determine its secondary stress pattern. In the middle of the two, there exists some variation and phonology-priority constraint ranking. Nevertheless, this normalized approach cannot generalize all of English secondary stress patterns, as exceptions still remain in the end. The exceptions will be marked with lexical-specific indexations.

**Keywords:** Secondary stress, Frequency, Base, Variation, Phonology-priority, Lexical-specific indexation

## 1. Introduction

English stress patterns are so complicated that there seems to be no complete account that is able to give a perfect explanation to them. Among such complexities, the diverse patterns of secondary stress are crucial. Even though we admit that primary stress may be listed in lexicon independently, the secondary stresses in derived words with the same suffix are hard to generalize in normalized grammar.

In quantitative studies so far, English secondary stress has been regarded as the result of interaction of two forces: stem-stress preservation (ID-Stress) and avoidance of stress clash (*Clash-Head). In this point, Optimality Theory which analyzes data with competition of constraints can give more improved explanation than rule-based theory.

Metrical stress by Generalized Alignment (McCarthy and Prince 1993), one of the OT-based explanations on stress pattern, however, is far from capturing lexical-idiosyncrasies.

As a solution to this, Pater (2000, 2006) proposes lexical-specific constraint ranking and morpheme-specific constraint indexation. Within his approach, all of exceptional stress patterns and lexical differences can be explained. It is hard, however, to believe that human memorize all of the lexical-specific constraint rankings.

Instead of jumping into lexical-specific analysis hastily, both Hammond (2003) and Cho (2004) argue the importance of frequency effects on English secondary stress pattern. Hammond claims that the more frequent the base form is, the more likely the derived form is to undergo reduction and this reduction has impact on secondary stress. Different form Hammond's approach, Cho (2004) suggested that constraint ranking be determined with comparison to maximum frequency of stem-stress preservation form.

This paper will make an attempt to prove that normalizing English secondary stress patterns should be approached separately according to its frequency based on corpus-based research. And the proposal on how to incorporate each frequency effect into grammar will be suggested.

## 2. Previous studies

### 2.1 Pater (2000, 2006)

Pater (2000) divides English secondary stress patterns in two groups (S1, S2)- the first group with stem-stress preservation, the second which do not preserve it - and gives each group different rankings of constraints.

In the first group, the demand of preserving stem- stress in derived forms makes stress clash. In other words, the constraint (1)Ident-Stress is more prominent than (2)*Clash-Head. In the second group, however, because of the benefit that stem-stress may not be preserved, derived forms do not bear any stress clash.

(1) Ident-Stress
  If a is stress, then f (a) must be stress

(2) * Clash- Head
NO stressed syllable may be adjacent to the head syllable of the Prosodic Word.

Related to his lexical-specific proposal, Pater (2006) develops the idea that each morpheme and constraint has its specific constraint indexation and they correlate with each other. One of the indexation-sensitive grammars is exemplified in (3) below. The unindexed stem shows variation (two optional candidates in the tableau), while the indexed stems show either mutation or deletion, depending on whether they are indexed to Max or Ident.

(3)

| Input | Output | *[ai]-L1 | Max-L2 | Ident-L3 | Max | Ident |
|---|---|---|---|---|---|---|
| /itara-i$_{L1}$-sa/ | itaraissa | *! | | | | |
| | ☞itarissa | | | | * | |
| | ☞itaroissa | | | | | * |
| /tavara$_{L2}$-i$_{L1}$-ssa/ | tavaraissa | *! | | | | |
| | tavarissa | | *! | | * | |
| | ☞tavaroissa | | | | | * |
| /jumala$_{L3}$-i$_{L1}$-ssa/ | jumalaissa | *! | | | | |
| | ☞jumalissa | | | | * | |
| | jumaloissa | | | *! | | * |

(8) presents the advantage of Pater's (2006) sketches in that it can easily capture variation(*itarissa, itaroissa*) and exception(*tavaroissa*). What remains doubtful is that whether human can memorize and learn all the morphemes with its specific indexation. As his proposal gives us a lot of burden of acquisition and memorizing process, the need of the more generalized approach arises.

Besides, lexical-specific approach cannot give any answer to the new words. That is, if new input is given, we cannot determine whether it has to be listed in S1 or S2. The claim by Pater, therefore, lacks productivity.

## 2.2 Hammond (2003)

Hammond(2003) explains the pronunciation of derived forms, assuming lexical frequency of the stem can affect the stress pattern. Fidelholtz(1975) has similar point of view with Hammond; frequency affects reduction of initial heavy syllables. Motivated by Fidelholtz's idea, Hammond argues that the more frequent the base form is, the more likely the derived form is to undergo reduction.

As he indicates the problems of Pater's (2000) approach, he gives a solution to explain the difference of words in S1 and S2 suggested above. If we list high-frequency words in the markedness constrain *Clash-Head, low-frequency word vacuously satisfies the constraint, as shown in (4).

(4)

| /transformation/ | C-H(···) | I-S |
|---|---|---|
| ☞[trásfor][màtion | | * |
| [tràns][fòr][má]tion | *! | |
| /exaltation/ | C-H(...) | I-S |
| [é][xá][tátion | *! | |
| ☞ [éxal][tátion | | |

Though this approach connects frequency effects and stress preservation pattern implicitly, the relation of the two is unnatural. The word which is frequently used indicates its unmarkedness in human knowledge system. Hammond makes unmarked words dominated by markedness constraint C-H(···). As he also suggested that relation of frequency and vowel reduction is 'inverse function of frequency', his approach cannot encode the frequency effect within the grammar.

## 2.3 Cho (2004)

Different from Hammond(2003), Cho (2004) thinks that every derived word might be listed in our lexicon independent of their base forms. In encoding frequency effect into grammar, she suggests new constraint (5) ID-stress ( f < t) to facilitate frequency into stem-stress preservation system.

(5) ID-stress ( f < t)
Preserve stem-stress if the frequency x of the input is smaller than t (threshold)
(Threshold is the minimum frequency among derived forms that stem-stress may not be preserved.)

In her data, derived form *creativity* ranks the highest frequency among stem-stress preservation group, so the threshold is based upon the frequency of this word: f=0.12. Therefore, we can make an example tableau like (6).

(6)

| /transformation/f=0.28 | ID-stress (f< 0.12) | *Clash-Head | ID-stress |
|---|---|---|---|
| ☞[trásfor][màtion | N/A | | * |
| [tràns][fòr][má]tion | N/A | *! | |
| /exaltation/f=0.01 | ID-stress (f< 0.12) | *Clash-Head | ID-stress |
| ☞[éxal][tátion | *! | | * |
| [é][xá][tátion | | * | |

In this analysis, threshold functions as the standard to both S1 and S2 group. Though she picks threshold among stem-stress preservation group, corpus-based research will show that stress preservation is much less frequent than no preservation, which indicates markedness nature of this phenomenon. It is natural to take the threshold as the most unmarked form, that is, the most frequent one to activate as a standard for all derivation process.

This paper will try to show the previous works' assumption that frequency affects English secondary stress pattern. The effect of frequency , however, should be encoded separately into grammar according to their patterns.

## 3. Research and proposal[1)]

### 3.1 English secondary stress

Copious amount of derived words which do not preserve stem-stress were found in the Brown and Frown Corpora. The frequencies of stem-stress not preserved group (henceforth NP) and preserved group (henceforth P) are summarized in (7) and (8) below.

(7) Stem-stress NOT preserved
Infórm 21 - Ínformàtion 542
Transfórm 23 - Trànsformátion 56
Transpórt 38 - Trànsportátion 97
Ségment 32 - Sègmentátion 1
Consúlt 28 - Cònsultátion 20
Consérve 10 - Cònservátion 60
Convérse 5 - Cònversátion 109
Confírm 37 - Cònfirmátion 14
Lamént 5 - Làmentátion 1
Phonétic 2 - Phònetícian 0
Cosmétic 9 - Còsmetícian 0
Average frequency of derived forms = 81.82

(8) Stem-stress preserved
Commúnal 16 - Còmmùnáliity 0
Cónglobate 0 - Cònglòbátion 0
Denóte 19 - Dènòtátion 0
Exhúme 0 - èxhùmátion 0
Condénse 2 - Còndènsátion 8
Cóntest 49 - Còntèstátion 0

1) Statistical data will be based upon quantitative analyses of the Brown Corpus and Frown Corpus . The data was mined by Wordsmith program. For this purpose the sample data were replicated from Pater (2000) and more specific classification was done with them.

Incrúst 0 - ìncrùstátion 0
Infést 1 - ìnfèstátion 2
Creáte166 - Crèàtívity 24
Exclúsive 50 - èxclùsívity 1
Immóbile 2 - ìmmòbílity 4
Doméstic 159 - Dòmèstícity 4
Elástic 8 - èlàstícity 5
Average frequency of derived forms = 3.43

Results in (6) and (7) present a couple of remarkable insights on the relation of frequency and stem-stress preservation. First of all, it can be seen that average frequency in (6) (81.82) is much higher than that of (7) (3.43). This might be partly because NP is regarded more unmarked in English stress system. Furthermore, there are no close ties between the frequency of base form and stem-stress. The statistic average is given in (8). Average here is calculated by means of dividing each frequency into total number of words.[2)]

(8) Frequency Average

| | Stem-stress NOT preserved | Stem-stress preserved |
|---|---|---|
| Base form | 0.01 | 0.02 |
| Derived form | 0.41 (Max = 2.71) | 0.017 (Max = 0.12) |

If this small sample data can give any implication, suggestions in Hammond 2003) and Cho (2004) should be revised in some respects. Admitting that the frequency of base form does not give direct indication to derived forms, Hammond's approach will hard to be supported. Besides, frequencies of derived forms in group NP marked much higher frequency, which indicates its unmarkeness. Choosing threshold as the most unmarked form in group P by Cho, therefore, is somewhat unnatural.

Proposal in this paper is that frequency average in (8) can give a crucial implication to secondary stress preservation pattern; if we hear some

2) 1 million in Brown + 1 million in Frown Corpus

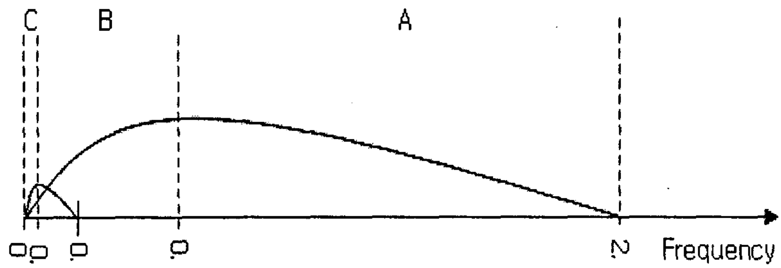
words frequently, we could be more familiar to them. Then there is no need of depending upon the base form for selecting its stress pattern. On the other hand, the unfamiliar words, which indicate low-frequency, want something to rely on for patterning its stress. The details will be discussed in section 3.2 below.

### 3.2 Effects of frequency in applying OT grammar

This section will show how frequency effects are encoded in OT grammar systematically. The statistical results from research in 3.1 can be summarized in graph (9). As grouping N and NP has relation only with derived forms, the frequencies of derived words are calculated regardless of its base. Group N and NP are calculated separately. In group NP, information marked the highest frequency, 542 among total data 20000. The value is calculated then 542/20000 *100 = 2.71. Now that 2.71 is the highest in Group NP, it marks in Max of the Group NP. As there exist words with no frequency in Group NP, the Min ranked in value 0. Since the same calculating method can be applied in Group P, Max of that group is based on the frequency of *creativity*, which marked the highest frequency 24. The value of Min is also 0, for the research shows 6 kinds of no frequency words in sample data.

The value of average is also normalized, since it is critical for evaluation of mid-frequency group. The detail will be discussed in section 3.2.3. For marking the average value, the total frequencies of derived forms in each group NP and P are summed up and then they are divided into the number of the derived forms-81.82 and 3.43 each. Then it was again divided with total size of the Corpus 20000. The calculation is then (81.82 / 20000)*100 = 0.41 for group NP and (3.43 / 20000)*100 = 0.017 for group P. Following is the graph showing the frequency distribution and the statistical results.

(9)




As the graph shows, words which do not preserve stem-stress are more frequent and they are much widely distributed in English words. For convenience, let us call the approach which argues the dependence of derived form on its base 'Source-oriented approach', and the one which compares derived forms independent of their base 'Product-oriented approach'. The source-oriented approach, then, does not give implication for the secondary stress pattern of high-frequency words. That is, frequently used words become so familiar as to have nothing to do with the stress pattern of its base in high-frequency words.

Another supporting evidence that source-oriented approach is not sufficient is that the derived forms from the same base shows different patterns. Derivations of inform and create are considered as the samples of the discussion.[3] The results are suggested in (10). Level classification has its basis on Borowsky (1993).

(10)
Level 1 suffixes
-al, -ry, -ous, -ic, -ical, -ate, -est, -ard, -ation, -al ,-ature, -atic -ive
Level 2 suffixes
-ing, -er, -ly, -y

With Level 1 suffix

Infórmal (P)
Ínformátion (NP)
Créature (NP)

3) As the purpose of this research is to show different stress preservation type of derived forms, exact frequency of each derived forms are not calculated.

Créátive (P)

Even though the suffixes are sorted in the same level, the stress preservation patterns of them are different. The discussion here indicates that English secondary stress pattern cannot be explained by traditional classification of suffix or merely by source-oriented approach.

### 3.2.1 Group 1: high-frequency words

Since solely source-oriented approach has been down graded from the discussion above, other analysis should be considered. The assumption that familiar words could be listed in lexicon independently without the help of its base form is relevant here. As the high-frequency ensures the familiarity, words with familiarity can be listed in our lexicon independently. Then as in (11), pure syllable-demanding constraint *Clash-Head should proceed correspondence constrain ID-Stress which has relation with stress pattern of its base.

(11)

| /information/ | *Clash-Head | ID-Stress |
|---|---|---|
| a.☞[ìnfor][mátion | | * |
| b. [ìn][fò[mátion | *! | |

If a new input with high-frequency falls into the range A in graph (9), the constraints hierarchy (11) will be applied. As the result of the constraint ranking *Clash-Head >> ID-Stress, *ìnformáion* will be a proper output, at the expense of the correspondence with its base stress.

### 3.2.2 Group 2: low-frequency words

Let us think about the process of determining English secondary stress, when the input in range C in graph (9) is provided. As the opportunity of hearing or using the form is rare, human might be hesitate about how to give stress pattern to that word. In this kind of situation, it is natural for human to seek something similar to depend on. It will be the base form that is similar to the derived forms, so this time base-dependent- source-oriented- stress patterning will be generalized. Therefore, ID-Stress will rank higher than *Clash-Head, which is a inversed process of Group 1.

(12)

| /condensation/ | ID-stress | *Clash-Head |
|---|---|---|
| a.☞[còn][dèn][sátion | | * |
| b. [cònden][sátion | *! | |

One crucial problem arises from this analysis. If derived forms can depend on base forms as its reliable threshold and the frequency effect also functions in this process, the frequency of all of the base forms should be very high. Among data in (8), there exist, however, some base forms which have low-frequency or even has no frequency-for example, *conglobate, infest,* and *exhume.*

Even though frequency of each word cannot ensure the preservation of base stress patter, the average of them can be shown as an evidence for source-oriented approach. Here, the average of base form frequency should be calculated for this purpose. The average of them in Group NP is normalized by dividing the total sum of the base frequency into the number of its kind; then 210/ 11 = 19.09. In the same way, the average in group P is 475/15 = 31.67. In spite of some words with much less frequency than the average frequency in group P, the average of group P is almost doubled than that of group NP. This statistic could be a piece of evidence that the tendency of relying on base forms could be more natural in group P than in group NP.

### 3.2.3 Group 3: mid- frequency

Copious amount of derived forms are in range B in graph (9). Range B is defined as the middle section between high and low frequency groups. Among the words which fall here, do they show either of the two patterns: variation or phonology-priority.

### 3.2.3.1 Secondary stress variation

None of previous studies discussed in section 2 give any analysis about the variably preserved words. The suggestion in this paper generally captures the phenomenon with maintaining the same constraints suggested for high and low- frequency words. The variation does not marked as an exceptional part of the stress patterns, but it can effectively incorporate with OT grammar as shown in (13).

(13)

| /augmentation/ | *Clash-Head | ID-Stress |
|---|---|---|
| a.☞[aù][mèn[tátion | * | |
| b.☞[aùment][tátion | | * |

As the frequency is not determined firmly, the constraints will have tie ranking. Of course, there is a possibility that one of the two constraints will be randomly chosen at the moment of speech. The standard of that choice is not just random but has its own criterion. Related analysis about this claim will be discussed in 3.2.3.2.

### 3.2.3.2 Competition of phonological demand and correspondence demand

Among data in range B are six examples which do not follow the frequency effects in its group. They are *conservation, transformation, consultation,* and *confirmation* in Group NP and *creativity* and *condensation,* in Group P. Words with suffix -*ician* and -*ity* are excluded, for they will be treated as lexical-specific-listing words in 3.2.4.1. Examples are lower than the average value of Group NP but higher than that of Group P. Those are defined as Mid-frequency group. Four among six satisfy the constraint ranking *Clash-Head >> ID-Stress and only two can be suited in constraint ranking ID-Stress >> *Clash-Head.

### 3.2.4 Exceptions - lexical-specific listing

In spite of the efforts of encoding frequency effect in generalized OT grammar, there still remains some exceptions which could not be captured in this normalized grammar. Provided that there is no way of encoding them into the grammar in relation with frequency effects, they should be treated as 'exceptions'.

Classifying data like *phonetician, cosmetician, creativity, elasticity* and *immobility* as exceptions is somewhat different from Pater (2000). He encodes all the lexical idiosyncrasies into the grammar. The proposal in this paper, however, makes frequency effects heeded by general grammar, and then encode residues which fall behind the normalized explanation as exceptions. The way of treating exceptions follows Pater's (2006) lexical-indexation suggestion. (14) shows one example of activation of lexical-specific constrains.

(14) lexical-specific constraint >> general constraints

| /phonetician/ | *Clash-Head -*ician* | ID-Stress | *Clash-Head |
|---|---|---|---|
| a.☞[phòne][tícian | | * | |
| b.[phò[nè[tícian | *! | | * |

In (14), the correct output *phònetícian* is selected because it satisfies the highest ranked lexical-specific constraint *Clash-Head-*ician*. The same explanation holds in other exceptional cases.

The exceptions can hold in another point of view, for example, suffix-based approach. Based upon the fact that all the exceptional patterns has suffix *-ician* or *-ity*, the idea can arise that stress patterns can be encoded in grammar according to the kind of suffix. As the frequency effect alone cannot give whole implication to English secondary stress pattern, other perspectives of approach should be considered.

## 5. Conclusion

This paper tries to show that English stress pattern in derived forms has some degree of relation with frequency effects. Different from research on the literature, the idea is based on the grouping of words according to their frequencies. Since the high-frequency words are familiar to human, it can be listed in our lexicon independently. The secondary stress within that group is determined not relying on their base forms. Low-frequency, however, indicates unfamiliarity. Secondary stress pattern of unfamiliar words has to have something to depend on, and the derived forms will select their base stress as the threshold. There also exist copious amount of variation, which fall into mid-frequency group. From the result that some mid-frequency words that do not show variation at the moment of pronunciation will follow phonological constraints at the expanse of correspondence requirement. This tendency suggests that stress patterns prefer to satisfy phonological demands prior to morphological ones. Given that there still remains some exceptions that cannot be encoded in this generalized frequency approach, the lexical-specific indexation activates at this moment.

## References

Cho, H. S. 2004. Frequency and stress preservation. *SNU Working papers in English Language and Linguistics* 3: 187-203.

Fidelholtz, J. 1975 Word frequency and vowel reduction in English *CLS* 11: 200-213

Hammand, M. 2003. Frequency, cyclicity and optimality. *Proceedings of the 2nd Korean International Conference on Phonology*. Seoul.

McCarthy, J. and A. Prince. 1993. *Generalized Alignment*. Ms. University of Messachusettes. Amherst, and Rutgers University.

Pater, J. 2000. Non-uniformity in English secondary stress: the role of ranked and lexically specific constraints. *Phonology* 17: 237-274.

Pater, J. 2006. The Locus of Exceptionality: Morpheme-Specific Phonology as Constraint Indexation. *ROA-866*.

Young-ah Do
youngahdo@gmail.com