An L2 Study on the Production of Stress Patterns in English Compounds*

Maria Luisa Zubizarreta, Xiao He, and Natalie Jonckheere

University of Southern California

This paper compares the production of stress patterns in various types of compounds by English natives and by L1 Spanish/L2 English speakers. English natives (ENs) systematically produced idiomatic compounds with fore-stress (Strong-Weak) patterns, and so did the second language learners (L2ers) (although to a lesser extent). Significant differences were found between the two groups in their stress production of less familiar compositional compounds. While there was a strong tendency for ENs to produce fore-stress patterns in the case of argument-head combinations and a weak tendency to produce end-stress (Weak-Strong) patterns in the case of modifier-head combinations, L2ers exhibited a strong tendency to produce end-stress patterns across the board. We attribute this to the end-based stress property of the L1, in conjunction with the relatively ambiguous prosody of the modifier-head type due to analogical processes.

Keywords:
Second Language Acquisition, stress patterns, compound words, noun-noun constructions, Nuclear Stress Rule

*We thank the USC Undergraduate Research Associates Program for funding our undergraduate RAs: Amitha Ganti, Rosanne Posslenzny, Tierney McBride, Natalie Jonckheere. We also thank Patricia Schneider-Zioga for assistance with data collection, and Chad Vicenik for the ToBI
tutorial in June 2010, where much feedback on our data was provided to us. We also thank Jason Rothman for invaluable editorial help.

1. Introduction

In this paper, we focus on one particular property of compounds, namely their stress patterns and the general algorithm that underlies them. We present a study on the production of stress patterns by English natives and by L1 Spanish/L2 English speakers in various types of compounds. We compare the results of the two populations and address the question of whether the second language learners (L2ers) are capable of acquiring the general algorithm that underlies the English compound stress patterns.

The paper is organized as follows. In section 2, we provide some background on the grammatical analysis of compounds. In section 3, we present our study: the research questions, the experimental design, participants’ characteristics, coding and results and the statistical analyses. In section 4, we discuss the results obtained, and in section 5, we conclude and point to future research venues.

2. Background

In his discussion of the stress criterion in English noun-noun constructions, Giegerich (2004, 2006) argues that all fore-stressed NNs (stress on the first N) in English must be lexical because
phrasal stress in English is typically final.\textsuperscript{1} This fore-stress nature of lexical compounds is attributed to the generalization put forth by Liberman & Prince (1977), which applies both to non-compound and compound words:

\begin{enumerate}
\item In any pair of sister nodes $[AB]_L$, where $L$ is a lexical category, $B$ is strong if it branches.
\item Fore-stress patterns are indeed typical of semantically non-compositional compounds, such as those in (2):
\begin{itemize}
\item bláckbird, blúeberry, gréenfinch, whitefly, gréyhound, rédworm, silverfish, pickpocket
\end{itemize}
\end{enumerate}

Yet, fore-stress is also found in semantically transparent compounds, in particular in cases where $N_1$ is unequivocally an “argument” of a deverbal head noun. Examples cited by Giegerich (2004) are given in (3) below, ranging from more to less familiar compounds. Note that while the argument often corresponds to a direct object (as in the examples below), it may also correspond to a locative PP complement, as in church-goer (one who goes to church) or roof-mounted antenna (antenna mounted on a roof).

\begin{enumerate}
\item a.wácht-maker, cóach-driver, sóap-dispenser, hédge-inspector, rásberry cook, síntax-provider, póetry generator.
\end{enumerate}

\textsuperscript{1} There is a long-standing debate in the literature as to whether compounds are lexical or phrasal. We do not directly address this debate here, but see Giegerich op.cit. for a review of this literature.
Giegerich contrasts the above type of compounds with those in which the first N bears an attributive relation (i.e. the “is a” relation) with respect to the head N, where “made of”, “belonging to” or “associated with” are considered to be particular instantiations of the attributive relation. These specific relations are inferred based on the properties of the attribute and on world knowledge. For example, the made-of relation is inferred from the attribute in *silk shirt* and *metal bridge* (shirt made of silk and bridge made of metal, respectively). Time and place references give rise to the associative relation, as in *village shop* (shop associated with village) and *morning coffee* (coffee associated with morning). The locative meaning “in the village” and temporal meaning “in the morning” are automatically inferred from the fact that “village” denotes a place and “morning” a time-reference, respectively.

While there is a tendency for attributive-relation to give rise to end-stress in N-N compounds (stress on the second N), it appears that it is a weaker tendency than fore-stress in the case of compound that involves an argument-relation. Giegerich (2004) remarks that end-stress in the former type of compounds is declining due to lexicalization and analogy. This author also notes that there is a certain degree of variability not only across, but also within, token items, and attributes this variability to different degrees of lexicalization: “End-stress in *steel bridge*, variable stress in *orange squash*, and fore-stress in *orange juice* are probably due to different degrees of diachronic stage of lexicalization.”

As is well-known, there are many minimal pairs of N-N compounds in English that are semantically distinguished by their stress-pattern (e.g. Giegerich op.cit.; Selkirk 1984; Zwicky 1986 among others). In the case of fore-stress, the underlying relation is an arg(ument)–head relation, as in the examples in (4). In the case of end-stress, as in (5), the relation between N1 and the head N2 is an attributive relation.
Note that the argument versus attributive distinction based on prosody also extends to adjectival compounds (examples below from Selkirk 1984):

(6) a. fróst bitten, diséase prone, blóod thirsty, gérm resistant
    b. lily white, waist hígh, dirt chéap, crytal cléar

An alternative view to the lexicalist approach to compound stress is offered by Cinque (1993). While acknowledging that the argument versus attributive semantic relations are at the core of the distinction between fore-stress and end-stress in compounds, this author proposes making a connection between the stress pattern found in compounds and in phrases in Germanic languages. In other words, the same Nuclear Stress Rule (NSR) that applies to phrases applies to compound structures as well. This author proposes to syntactically encode arguments within a compound as a complement relation via the “sister-of” relation and the attributive (or modifier)
relation as a “specifier-of” relation. Cinque furthermore proposes that the most deeply embedded node (computed in terms of dominating nodes) attracts NS. The head being more deeply embedded than its specifier, the head attracts NS in the case of the modifier-head relation, as exemplified in (7a). On the other hand, the complement being more deeply embedded than the head, the complement attracts NS in an arg-head relation, as exemplified in (7b).²

(7) a. [kitchen [Ø ráck]]
   b. [kitchen [[tòwel] rack]]

Cinque thus captures the similarity between compound stress in examples such as (7) and phrasal stress in OV/VO structures in Germanic languages, where the object (defined as sister of the head) systematically attracts NS, unlike modifiers (defined as specifiers of the head). It also accounts for the semantics associated with the variable stress patterns in compounds such as (3) and (4), as well as for the semantics associated with the variable phrasal stress patterns in the German examples below (examples from Krifka 1984). In (8a) the PP argument/complement attracts NS, while the PP attribute/modifier in (8b) does not and NS goes on the verb instead.³

(8) a. Hans hat an seinem Papier gearbeitet.

Hans has on his paper worked

² Cinque gives no independent justification for the null complement in examples such as (7a).
³ Note that contrary to standard assumptions, even English has cases of non-final phrasal stress, namely in eventive SV structures, such as the examples below (see Selkirk 1984; Zubizarreta 1998; Zubizarreta & Vergnaud 2005 and references cited therein). This lends further support to Cinque’s view that there is more in common between compound stress and phrasal stress than commonly assumed.
(i) The máil arrived.
(ii) We saw [a whále swimming].
(iii) There’s [a fláw in the argument].
‘Hans has worked on his paper.’

b. Hans hat in seinem Büro gearbeitet.

Hans has in his office worked

‘Hans has worked in his office.’

In this paper, we will focus exclusively on compound stress. Whether one adopts a structural account (as Cinque op.cit) or a semantic account (as Giegerich op.cit.; Selkirk 1984 and others) to define the relation between the two lexical components of a compound word, the following generalization regarding English compound stress seems to emerge:4

(9) a. Lexicalized compounds (as in semantically opaque compounds) have fore-stress.

b. In non-lexicalized compounds,

(i) arg-head compounds have fore-stress

(ii) mod-head compounds have end-stress, but lexicalization and analogy can give rise to fore-stress patterns.

---

4 Plag, Kunter, Braun & Lappe 2008 conducted a study on the perception of compound stress by native speakers of American English. A large variety of compounds were extracted from the Boston Corpus and rated for prominence on a continuous scale (by moving a slider on a computer screen), with ratings being later transformed into values from 0 through 999 (with higher numbers indicating more rightward stress). Based on this study, the authors concluded that stress placement is not deterministic but probabilistic and that argument structure is not a relevant factor. They conclude that lexicalization (as determined by spelling and to a lesser extent by frequency) is a relevant factor, as well as the semantic type (whether the compound refers to an individual or instrument versus an action or result). This is not the place for a review of this article, which raises many complex issues, but one point to note is the type of coding used, which may have skewed the results. Also, given our own experience with coding, it is a fatal mistake to code out of context for prominence of material produced as part of a larger discourse. We have found that coders often need the larger intonational context to make a secure judgment, and particularly important is whether the compound occurs sentence internally or not (i.e. whether or not it occurs at the edge of an intonational break). Equally confusing is the request that the coder place the slider in the middle if equally stressed, giving rise to an ambiguity between genuine equal stresses and “not sure”. Furthermore, it is not how much stronger/weaker A is perceived with respect to B, but rather to what extent native speakers can agree whether A is stronger/weaker than B.
3. Our study

The design of this study includes three types of compounds, which aim to capture two types of properties: (1) the two ends of the spectrum regarding lexicalization as defined in terms of compositionality: semantically opaque compounds and semantically transparent compounds, and (2) the argument-modifier distinction. As mentioned earlier, frequency is also an important consideration, such that compositional compounds that are very frequent might be committed to memory as such and retrieved without going through a compositional analysis. Since there is no frequency table for compounds readily available, we decided to capture frequency in terms of familiarity.

The study was designed to address the following questions regarding the acquisition of English stress patterns by L1 Spanish/L2 English learners. This L2 population is of particular interest because Spanish has systematically end-stress in compounds as well as in phrases (see Zubizarreta & Nava 2011 and references cited therein). In Spanish, there is the confound that compounds are systematically head-initial; word order cannot be distinguished from the type of grammatical relation that the components of the compounds entertain. For example, corta-papel ‘paper cutter’ (head-arg), azul cielo ‘sky blue’ (head-mod).\(^5\)

---

\(^5\) The role of head-directionality in the L2 acquisition of English compounds merits investigation and we hope to address this in future research.
a. Are L2ers more native-like in their stress production of frequent/idiomatic N-N compounds than in their stress production of infrequent/compositional N-N compounds?

b. More specifically, would L2ers produce significantly more fore-stress in the case of frequent/idiomatic compounds than in the case of infrequent/semantically-transparent arg-head compounds?

c. In the case of infrequent/semantically-transparent compounds, would L2ers make a distinction between arg-head and mod-head compounds, producing more fore-stress in the former type of compound than in the latter?

d. What extra-linguistic factors, if any: Age of Arrival (A0A), Age of Exposure (AoE), Length of Residency (LoR), are relevant in the production of native-like prosodic patterns in English compounds?

3.1 Materials and Methods

The task designed to gauge the stress pattern in compounds was a reading protocol that consisted of short passages containing the target compounds. The passages were taken mostly from the journal *The Economist* and the worldwide web, simplified and shortened where needed. In a pilot study, 20 native speakers were asked to rank the compounds on a scale of 1 (least familiar) to 10 (most familiar). A subset of the passages was selected based on the familiarity results and overall fluency with which native speakers read the passages.

Three types of compounds were targeted: one was a set of lexicalized compounds, as determined by its orthography (written as one word) and its non-compositional (idiomatic)
meaning; the other was a set of non-lexicalized compounds: semantically transparent and orthographically written as two words. Among the latter, there were two subsets: one set of compounds consisted of the arg-head relation and the other consisted of the mod-head relation. For the final protocol, a total of 24 compounds were selected: 8 idiomatic, highly familiar compounds and 16 compositional, unfamiliar or less familiar compounds (8 consisting of a mod-head relation and 8 consisting of an arg-head relation). The compounds were 2 or 3 syllables long, with the exception of the 4-syllable long sugar-selling. The arg-head and the mod-head set of compounds each contained 2 deverbal –er, 2 deverbal –ing, two adjectival –ed, and 2 bare N-N compounds.

(11) *Lexicalized compounds*: shopkeepers, highchair, headache, shoeboxes, toothpaste, goldfish, firefighter, blackboard.

*Non-lexicalized compounds*:

*Arg-head relation*: cheese-eaters (one who eats cheese), wage-earner (one who earns a wage), sugar-selling (one who sells sugar), bribe-taking (one who takes bribes), roof-mounted (mounted on roof), country-based (based on country), apple farm (farm that produces apples), fruit store (store that sells fruits)

*Mod-head relation*: noon dancing (dancing at noon), day sleeping (sleeping during the day), Harvard trained (trained at Harvard), city raised (raised in the city), village phones (phones in villages), toy pet (pet that is a toy), stone buildings (buildings made of stone), metal box (box made of metal).
The passages were randomized and presented in a power point format, with one passage per slide. For the full list of passages, see the Appendix. The reading passages were preceded by a warm up of 6 short sentences. Participants were asked to read at a natural pace but not to rush; the reading was recorded with the Praat software for later analysis.

The assumption underlying the familiarity distinction was that lexicalized, familiar compounds were possibly committed to memory along with their fore-stress patterns, but not the semantically transparent, unfamiliar ones. Therefore, we do not expect fluctuation in native speakers’ judgments with the lexicalized ones. On the other hand, if Giegerich is correct (see section 2), in the case of semantically transparent, unfamiliar compounds, we expected to see more stability in the production of fore-stress in the case of arg-head compounds, and possibly less stability in the production of end-stress patterns in the case of mod-head compounds.

As for L2ers, we would expect that in the case of lexicalized compounds, “familiarity” would be a predictor for fore-stress pattern production, given our hypothesis that more familiar compounds would be stored in memory with its prosodic pattern. On the other hand, in the case of semantically compositional, unfamiliar compounds, there are a priori two possible scenarios: (1) L2ers’s production is native-like, (i.e. more fore-stress in arg-head compounds and more end-stress in mod-head compounds), (2) L2ers tend to produce end-stress more than fore-stress across the board, due to the influence of their L1. In the case of L2ers, we also aim to determine whether non-linguistic variables, such as AoA (Age of Arrival), AoE (Age of Exposure), LoR (length of residence in the US), are relevant in L2ers production of fore-stress patterns.

3.2 Participants
The study included 16 native speakers of English, who were undergraduate students at the University of Southern California (USC) at the time of testing, and 16 L1 Spanish / L2 English speakers of English, who were undergraduate or graduate students at USC or California State University-Fullerton or who were Spanish lecturers or professors at USC or nearby institutions. The L2ers were from Spain, Mexico, and Peru. The age range of the native speakers at testing was 18-22 (mean age 20). The age range of the L2ers at testing was 18-48 (mean age 31.6). For the L2ers, the age range of AoA was 7-44 (mean AoA 19.5), the age range of AoE was 3-17 (mean AoE 9.6), and the range of LoR in the US was 2-23 (mean LoR 11.6).

3.3 Coding and results

The recordings were split and coded by two Undergraduate Assistants based on perception. The results obtained by the two coders were then compared; the inter-rater reliability based on these initial coding was computed and is reported in the table below. All discrepancies were judged by a third undergraduate RA and then discussed in lab meetings to reach a consensus. The ones that remained controversial were later examined at an intonational hands-on workshop lead by a ToBI expert (at USC on June 2010). The final results are summarized in the two tables below. Compounds with fore-stress are tallied under W1, those with end-stress are tallied under W2, those that remained indeterminate were tallied under the question mark sign, and those that were not pronounced as a compound (i.e. with a slight pause between the two words) are tallied under NC.

---

6 The sum of percentages along the “total” row may not equal 100% due to rounding.
As we can see from the familiarity ratings, both the native controls and the L2ers considered the idiomatic compounds as far more familiar than the non-idiomatic ones. We can also note that, as expected, the idiomatic ones were without exception pronounced with fore-stress by the native speakers. Although the picture was not as clear-cut as with idiomatic compounds, there was a strong tendency for natives to pronounce the arg-head compounds with fore-stress. As for the mod-head compounds, while there was an overall tendency to pronounce this type of compound with end-stress, there was a fair degree of variability across and within token items.

Table 1. English Natives

<table>
<thead>
<tr>
<th>Argument-Head</th>
<th>W1</th>
<th>W2</th>
<th>?</th>
<th>Mean Familiarity</th>
<th>Inter-rater Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cheese eaters</td>
<td>15 (93.75%)</td>
<td>0</td>
<td>1 (6.25%)</td>
<td>3.000</td>
<td>93.75%</td>
</tr>
<tr>
<td>Wage earner</td>
<td>14 (87.50%)</td>
<td>1 (6.25%)</td>
<td>1 (6.25%)</td>
<td>3.670</td>
<td>93.75%</td>
</tr>
<tr>
<td>Sugar selling</td>
<td>16 (100%)</td>
<td>0</td>
<td>0</td>
<td>3.500</td>
<td>100%</td>
</tr>
<tr>
<td>Bribe taking</td>
<td>15 (93.75%)</td>
<td>1 (6.25%)</td>
<td>0</td>
<td>6.670</td>
<td>93.75%</td>
</tr>
<tr>
<td>Roof mounted</td>
<td>13 (81.25%)</td>
<td>2 (12.50%)</td>
<td>1 (6.25%)</td>
<td>4.500</td>
<td>93.75%</td>
</tr>
<tr>
<td>Country based</td>
<td>10 (62.50%)</td>
<td>4 (25%)</td>
<td>2 (12.50%)</td>
<td>3.170</td>
<td>68.75%</td>
</tr>
<tr>
<td>Apple farm</td>
<td>14 (87.50%)</td>
<td>2 (12.50%)</td>
<td>0</td>
<td>4.330</td>
<td>81.25%</td>
</tr>
<tr>
<td>Fruit store</td>
<td>15 (93.75%)</td>
<td>0</td>
<td>1 (6.25%)</td>
<td>3.500</td>
<td>93.75%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>112 (87.50%)</strong></td>
<td><strong>10 (7.81%)</strong></td>
<td><strong>6 (4.69%)</strong></td>
<td><strong>4.043</strong></td>
<td><strong>89.84%</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Modifier-Head</th>
<th>W1</th>
<th>W2</th>
<th>?</th>
<th>Mean Familiarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noon dancing</td>
<td>0</td>
<td>16 (100%)</td>
<td>0</td>
<td>2.000</td>
</tr>
<tr>
<td>Day sleeping</td>
<td>10 (62.50%)</td>
<td>6 (37.50%)</td>
<td>0</td>
<td>2.500</td>
</tr>
</tbody>
</table>
As for the L2ers, there was a higher degree of indeterminacy in the coding of their data than in the coding of the native speakers and 27 renditions were pronounced as non-compounds (NC); these token items had some degree of pause between W1 and W2. (More of the NC type is observed in the arg-head cases, less so in the mod-head cases, and hardly any in the idiomatic, frequent compounds. Note, on the other hand, that native speakers never produced compounds with a pause between them). Putting aside the NC case, there was a strong tendency for L2ers to
pronounce the idiomatic compounds with fore-stress, while the stress placement for the non-
idiomatic compounds was highly variable, but with more end-stress than fore-stress production
overall.

### Table 2. L2ers

<table>
<thead>
<tr>
<th>Argument-Head</th>
<th>W1</th>
<th>W2</th>
<th>2</th>
<th>NC</th>
<th>Mean</th>
<th>Inter-rater Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W1</td>
<td>W2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheese eaters</td>
<td>2 (12.50%)</td>
<td>12 (75%)</td>
<td>2 (12.50%)</td>
<td>0</td>
<td>3.813</td>
<td>81.25%</td>
</tr>
<tr>
<td>Wage earner</td>
<td>3 (18.75%)</td>
<td>9 (56.25%)</td>
<td>1 (6.25%)</td>
<td>3 (18.75%)</td>
<td>3.625</td>
<td>81.25%</td>
</tr>
<tr>
<td>Sugar selling</td>
<td>0</td>
<td>12 (75%)</td>
<td>0</td>
<td>4 (25%)</td>
<td>2.375</td>
<td>68.75%</td>
</tr>
<tr>
<td>Bribe taking</td>
<td>3 (18.75%)</td>
<td>11 (68.75%)</td>
<td>2 (12.50%)</td>
<td>0</td>
<td>2.688</td>
<td>87.50%</td>
</tr>
<tr>
<td>Roof mounted</td>
<td>5 (31.25%)</td>
<td>6 (37.50%)</td>
<td>1 (6.25%)</td>
<td>4 (25%)</td>
<td>3.563</td>
<td>50.00%</td>
</tr>
<tr>
<td>Country based</td>
<td>8 (50%)</td>
<td>6 (37.50%)</td>
<td>0</td>
<td>2 (12.50%)</td>
<td>4.000</td>
<td>62.50%</td>
</tr>
<tr>
<td>Apple farm</td>
<td>8 (50%)</td>
<td>5 (31.25%)</td>
<td>0</td>
<td>3 (18.75%)</td>
<td>3.688</td>
<td>62.50%</td>
</tr>
<tr>
<td>Fruit store</td>
<td>4 (25%)</td>
<td>6 (37.50%)</td>
<td>0</td>
<td>6 (37.50%)</td>
<td>6.188</td>
<td>62.50%</td>
</tr>
<tr>
<td>Total</td>
<td>33 (25.78%)</td>
<td>67 (52.34%)</td>
<td>6 (4.69%)</td>
<td>22 (17.19%)</td>
<td>3.742</td>
<td>69.53%</td>
</tr>
<tr>
<td>Modifier-Head</td>
<td>Noon dancing</td>
<td>0</td>
<td>15 (93.75%)</td>
<td>0</td>
<td>2.563</td>
<td>93.75%</td>
</tr>
<tr>
<td></td>
<td>Day sleeping</td>
<td>3 (18.75%)</td>
<td>9 (56.25%)</td>
<td>1 (6.25%)</td>
<td>4.188</td>
<td>62.50%</td>
</tr>
<tr>
<td></td>
<td>Harvard trained</td>
<td>13 (81.25%)</td>
<td>1 (6.25%)</td>
<td>1 (6.25%)</td>
<td>4.813</td>
<td>62.50%</td>
</tr>
<tr>
<td></td>
<td>City raised</td>
<td>4 (25%)</td>
<td>9 (56.25%)</td>
<td>0</td>
<td>3 (18.75%)</td>
<td>4.813</td>
</tr>
<tr>
<td></td>
<td>Village phones</td>
<td>6 (37.5%)</td>
<td>8 (50%)</td>
<td>2 (12.50%)</td>
<td>1.563</td>
<td>87.50%</td>
</tr>
<tr>
<td></td>
<td>Toy pet</td>
<td>1 (6.25%)</td>
<td>14 (87.50%)</td>
<td>0</td>
<td>1 (6.25%)</td>
<td>4.125</td>
</tr>
<tr>
<td></td>
<td>Stone buildings</td>
<td>0</td>
<td>16 (100%)</td>
<td>0</td>
<td>0</td>
<td>4.688</td>
</tr>
<tr>
<td></td>
<td>Metal box</td>
<td>2 (12.50%)</td>
<td>12 (75%)</td>
<td>0</td>
<td>2 (12.50%)</td>
<td>5.063</td>
</tr>
<tr>
<td>Total</td>
<td>29 (22.66%)</td>
<td>84 (65.63%)</td>
<td>4 (3.13%)</td>
<td>11 (8.59%)</td>
<td>3.977</td>
<td>76.56%</td>
</tr>
</tbody>
</table>
We examined the L2 data by subject and found that, in the case of idiomatic compounds, five participants had a native (8/8) and three participants had a quasi-native (7/8) fore-stress production. On the other hand, for the case of arg-head compounds, only one participant produced over 50% of fore-stress patterns (5/8). We turn next to the statistical analysis of the data.

### 4. Statistical Analysis and results

All statistical analyses were conducted in the statistical computing program **R**. For the statistical analyses, we excluded one of the L2ers, whose speech was not fluent (it had lots of hesitations). Therefore, the analyses presented below are based on 16 native speakers and 15 L2ers. In
addition, we excluded tokens that were deemed to be non-compounds or indeterminate by the raters. All the excluded tokens constitute approximately 8.9% of all tokens. The proportions of fore-stress placements for each compound type and each participant group after data exclusion is shown in the graph below.

**Figure 1. Proportion of fore-stress placement**

We are interested in (1) how L2ers’ fore-stress production compares with that of native speakers on each of the three compound types and (2) the comparison of fore-stress production within and across participant groups. In other words, we would like to know whether L2ers and native speakers differ, overall, with respect to fore-stress production, whether fore-stress production differed across types of compounds, and whether fore-stress placement on different
types of compounds was contingent on participants’ native language (English vs. Spanish). To this effect, we carried out analyses of variance (ANOVAs) on participants’ fore-stress placements by testing (a) a main effect of PARTICIPANT type (i.e. native speakers versus L2ers), (b) a main effect of COMPOUND types (i.e. arg-head versus mod-head versus idiomatic), and (c) an interaction effect of PARTICIPANT and COMPOUND. Because the stress placement data were binomial responses (i.e. presence versus absence of fore-stress), in order to conduct by-subject and by-item ANOVAs, we first calculated each subject’s percentage of fore-stresses on each compound type and each item’s percentage of fore-stresses within each participant group and then arcsine-transformed these percentages. An effect is considered significant only if it reaches significance ($p \leq .05$) on both by subject- and by item-tests. To examine whether the effects of non-linguistic (independent) variables, such as AoA, AoE, LoR, and familiarity, are relevant in L2ers production of fore-stress patterns, we conducted logit mixed-effects (LME) regressions on the raw binomial data (i.e. presence/absence of fore-stress) with subject- and item-specific random effects (Baayen 2008; Jaeger 2008).

4.1 Results of Analyses of Variance

We found a significant main effect of COMPOUND type ($F_1(2,58) = 130.795, p < .001; F_2(2,21) = 27.187, p < .001$), suggesting that fore-stress placement patterns vary across the three types of compounds. We also observed a significant main effect of PARTICIPANT ($F_1(1,29) = 94.134, p < .001; F_2(1,21) = 27.187, p < .001$), which suggests that overall, native speakers were significantly more likely than their non-native counterparts to produce fore-stress patterns. In addition, we found a significant interaction between COMPOUND and PARTICIPANT ($F_1(2,58) = 130.795, p < .001; F_2(2,21) = 27.187, p < .001$).
Due to this interaction, we conducted a series of planned comparisons to obtain more fine-grained analyses.

Let us first look at the stress production within each group across compound types. For native-speakers, we found that both idiomatic and arg-head compounds elicited significantly more fore-stress patterns than did mod-head compounds (idiomatic vs. modifier-head: \( p < .001 \); arg-head vs. mod-head: \( p < .001 \)). In addition, idiomatic compounds elicited significantly more fore-stress patterns than arg-head compounds (\( p < .05 \)). As for L2ers, we found that this group produced significantly more fore-stress patterns on idiomatic compounds than on the other two types of compounds (\( p < .0001 \)). However, arg-head and mod-head compounds did not differ significantly (\( p > .1 \)).

We turn next to the comparison between L2ers and native speakers’ production of fore-stress patterns. We found that compared to native speakers, L2ers produced significantly fewer fore-stress patterns on both arg-head compounds (\( p < .01 \)) and idiomatic compounds (\( p < .05 \)). However, the two groups of participants did not differ significantly on the mod-head compounds (\( p > .08 \)).

4.2 Regression Analyses: Effects of non-linguistic variables on fore-stress production

The next set of analyses examines whether non-linguistic variables, such as AoA, AoE, and LoR, and familiarity affect L2ers’ fore-stress placement in compounds. We used LME models in the ‘lme4’ package, (for extensive discussion, see Jaeger 2008; Gries 2009). We fit LME models for each compound type by including the four non-linguistic variables as fixed predictors and subject and item variables as random effects. For the arg-head and mod-head
compounds, none of the four non-linguistic variables was a significant factor in L2ers’ fore-stress production (see table 3 below). However, for the idiomatic compounds, both AoA and AoE were significant predictors of L2ers fore-stress production, but not LoR and familiarity. Specifically, there was an inverse relationship between the number of fore-stress patterns produced on the one hand, and AoA and AoE on the other hand, such that participants who had been exposed to English or had come to the US at a younger age were more likely to stress the first word of idiomatic compounds.

Table 3. Non-linguistic variables in production of fore-stress patterns

<table>
<thead>
<tr>
<th>Compound type</th>
<th>Variables</th>
<th>Standard Error</th>
<th>z value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argument-head</td>
<td>AoA</td>
<td>0.03504</td>
<td>0.613</td>
<td>p = .54</td>
</tr>
<tr>
<td></td>
<td>AoE</td>
<td>0.06481</td>
<td>0.242</td>
<td>p = .84</td>
</tr>
<tr>
<td></td>
<td>LoR</td>
<td>0.05692</td>
<td>-0.206</td>
<td>p = .95</td>
</tr>
<tr>
<td></td>
<td>Familiarity</td>
<td>0.11158</td>
<td>-0.063</td>
<td>p = .81</td>
</tr>
<tr>
<td>Modifier-head</td>
<td>AoA</td>
<td>0.04711</td>
<td>-0.571</td>
<td>p = .57</td>
</tr>
<tr>
<td></td>
<td>AoE</td>
<td>0.07711</td>
<td>-0.499</td>
<td>p = .62</td>
</tr>
<tr>
<td></td>
<td>LoR</td>
<td>0.06879</td>
<td>-0.059</td>
<td>p = .95</td>
</tr>
<tr>
<td></td>
<td>Familiarity</td>
<td>0.13589</td>
<td>-1.073</td>
<td>p = .28</td>
</tr>
<tr>
<td>Idiomatic</td>
<td>AoA</td>
<td>0.05314</td>
<td>-1.992</td>
<td>p &lt; .05*</td>
</tr>
<tr>
<td></td>
<td>AoE</td>
<td>0.11016</td>
<td>-1.965</td>
<td>p &lt; .05*</td>
</tr>
<tr>
<td></td>
<td>LoR</td>
<td>0.08122</td>
<td>1.570</td>
<td>p = .12</td>
</tr>
<tr>
<td></td>
<td>Familiarity</td>
<td>0.14234</td>
<td>-0.127</td>
<td>p = .90</td>
</tr>
</tbody>
</table>
5. General Discussion

To briefly recapitulate, lexicalized compounds in English are typically fore-stressed. On the other hand, in the case of non-lexicalized compounds, there is a general algorithm that determines the placement of main stress based on the semantic relation that the two parts of the compound entertains: in the case of an arg-head compounds, main stress goes on the argument, and in the case of mod-head compounds, main stress goes on the head, with a certain degree of variability due to lexicalization and analogy. The generalization established for lexicalized compounds was strongly verified by our study: All English speakers clearly produced fore-stress patterns 100% of the time in the case of lexicalized compounds. The generalization for non-lexicalized, arg-head compounds was verified to a great extent as well: English speakers produced fore-stress patterns 92% of the time (among those renditions that were judged to have clear or unambiguous prosody). The remaining 8% could be attributed to “noise” in the data, for example, to task-related errors.

As for the non-lexicalized, mod-head compounds, the results obtained by our study were puzzling. Of all the clear or unambiguous renditions, only 65% had end-stress patterns. The remaining 35% had fore-stress, although the compounds were rather low in the familiarity scale. As can be seen in Table 1, there is considerable variability in native speakers’ rendition of day sleeping (familiarity: 2.5), of city raised (familiarity: 3.8), and of village phone (familiarity: 1.3), and a particularly high convergence on fore-stress in the case of Harvard-trained. We may speculate that the grammatically-determined final stress pattern of a Modifier + N compound may be superseded in some cases by the fore-stress pattern of a relatively frequent/lexicalized compound due to analogy. The question then arises how analogy works for the online task at
hand: does the first N of the compound (i.e. the modifier) or the second N of the compound (i.e. the head) trigger the analogical process?

Work by Gagné & Shoben (1997) argues in favor of the view that it is the modifier that is relevant (at least in the case of an online, reading task). In their study on relational priming in N-N combinations, these authors showed that how likely a certain relation is used depends on the “relational distribution” of the modifier (i.e. the first N), where “relational distribution” reflects a person’s experience with that N in combination with other Ns. They furthermore demonstrated that the modifier’s relational frequency determines the ease with which a N-N combination can be interpreted. Furthermore, Gagné & Shoben crucially showed that the head’s relational frequency does not give rise to such facilitating effects.

It is then plausible to hypothesize that a mod-head compound that is relatively frequent could affect the (stress) production of another less frequent mod-head compound in an online reading task, where both compounds share the same modifier. Thus, it is possible that the final stress pattern of day sleeping (Google count: 646,000) may be superseded by the fore-stress pattern of more frequent/lexicalized compounds, such as daydreaming (Google count: 7,450,000). Similarly, city raised (Google count: 236,000) may be superseded by forms such as City Hall (Google count: 17,300,000) and Harvard-trained (Google count: 474,000) by forms such as Harvard University (18,500,000), where City Hall and Harvard University are relatively frequent/lexicalized compounds with fore-stress patterns.

As for the Spanish L1/English L2 group, our study shows that they are able to acquire fore-stress in idiomatic compounds. Of the clear, compound-like production, 75% of idiomatic compounds were produced with fore-stress. Although as a group, they were not native-like, individual data show that 8 out of the 15 fluent speakers were native-like (with 8/8 fore-stress
production) or near native-like (with 7/8 of fore-stress production). This shows that L1 Spanish/L2 English speakers do not have a problem producing Germanic trochaic stress (Strong-Weak) patterns per se, although AoA and AoE do seem to be a determinant factor. Indeed, we found that early age of exposure to English and early age of arrival in the US were significant predictors of L2ers’ fore-stress production, although length of LoE (length of exposure) was not. Unexpectedly, “familiarity” was not a significant predictor either. Although mean familiarity for idiomatic compounds (slightly over 7) was higher than for the other semantically, compositional idioms (slightly under 4), it did not reach statistical significance. This factor merits further investigation with a larger sample stimulus data and a larger population.

In the case of unfamiliar, compositional compounds, the L2ers produced a low amount of fore-stress patterns, regardless of AoA and AoE: 33% in the case of arg-head compounds and 25% in the case of mod-head compounds (among the clear, compound-like renditions), and only one L2 learner produced over 50% (5/8) of fore-stress pattern in the arg-head type compound. This goes to show that acquiring the stress distinction based on the relationship that the two parts of the compound entertain (argument versus modifier) is extremely hard to acquire for L1 Spanish/L2 learners of English, even for those with an early AoA or AoE. 7 L2ers tend to overwhelmingly produce the L1 (end-stress) patterns. In fact, L2ers and native speakers were statistically similar only in the case of the mod-head type of compound, which English natives also tended to produce with end-stress.

---

7 It is noteworthy that in the case of Harvard-trained (which belongs to the mod-head type), 13 L2ers produced patterns with fore-stress and 13 natives speakers did so too, possibly influenced, as suggested earlier, by the fore-stress pattern of Harvard University (a very frequent lexical item in academia). This reinforces the hypothesis that the cases of fore-stress produced by the L2ers are due to stress-pattern memorization in the case of frequent compounds, or in the case of infrequent ones, the fore-stress pattern is produced by analogy with a frequent/similar compound.
If the small sample of compounds in our study is in any way a reflection of native speakers’ production of compound stress, it is perhaps not surprising that it is challenging for L2ers to gauge the divide between arg-head and mod-head given that the input is extremely noisy in the case of mod-head compounds, due presumably to analogical processes. As mentioned earlier, 35% of native renditions in the case of mod-head compounds were produced with fore-stress and this would appear to blur the grammatical-based stress distinction for the L2 learner, although apparently not for the L1 learner. In other words, the level of noise in the data that is tolerated by L2 and L1 learners might be different, in particular when the prosodic patterns of the two languages in contact overlap. Indeed, in Spanish, main stress in compound, as in phrases, is systematically final. This means that in Spanish compounds, which are head initial, main stress falls on the second N regardless of the grammatical relation that the two Ns entertain. Faced with a significant degree of ambiguity in the input, the L2er adopts the prosodic pattern that coincides with that of the L1, namely end-based stress. The L1 learner, on the other hand, has no basis for such a bias, at least off-hand, and is therefore able to converge on the target grammar despite the variability in the input. This hypothesis could be formalized in a model like that proposed by Yang 2002, in which different grammars in competition are assigned different weights (depending on the input). It is reasonable to assume that in the case of L2 acquisition, the L1 starts out with a higher weight than the L2, and that it requires a significant amount of consistent data for the learner to converge on the target grammar. It is an empirical issue to determine what the tolerable threshold is with respect to input variability beyond which the L1 grammar cannot out-compete the L2 grammar.
6. Conclusion

To conclude, we have found that L1 Spanish/L2 English speakers can acquire trochaic stress patterns, and they do so readily in case of frequent/lexicalized compounds, although their success is contingent on AoA and AoE. Although we did not find that “familiarity” reached statistical significance as a predictor for fore-stress production in L2ers as a group, based on the comparatively high mean obtained for this class of compounds, we may nevertheless tentatively conclude that such forms are stored in memory with their associated fore-stress pattern. On the other hand, in the case of relatively unfamiliar compounds, there is a strong tendency for L2ers to produce end-stress patterns, irrespectively of AoA and AoE. We suggested that this is due to the relatively high degree of noise in the input regarding fore-stress vs. end-stress distribution, and that in such situations the L2 grammar cannot out-compete the L1 grammar, which dictates end-stress across the board.

The results of the present study are consistent with the findings reported in the study by Zubizarreta & Nava (2011). That study found that the Germanic trochaic pattern in intransitive (SV) sentences is very challenging for native speakers of Spanish to acquire. Of the 47 L2ers tested in that study, only 9 produced over 50% of Germanic stress patterns, and the amount produced was significantly lower than the amount produced by the native English speakers.

Finally, we would like to point out the limitations of this study. First, a larger amount of compounds with different types of morphological endings needs to be studied and a larger L2 population needs to be tested. Second, we yet need to understand the role of analogy with respect to prosody, not only in L2 production but also in the case of L1 production. Therefore this is an important question for future investigation. This could be investigated with a priming
experiment, in which the elicited target item is preceded by an oral presentation of analogical compounds that vary according to frequency and degree of lexicalization. Indeed, we would like to know whether priming of frequent analogical compounds can increase the production of fore-stress patterns in the case of unfamiliar compounds, both for native speakers as well as for L2ers. And finally, further investigation on prosodic variation would give us much needed insight into the role of input in L2 acquisition, in particular in determining what the tolerable threshold is for prosodic variation beyond which sub-regularities cannot be extracted from the input data by L2 learners. This could be investigated by teaching compounds with an English syntax/prosody and with an artificial lexicon (along the lines of current work by E. Newport), so as to factor out the role of analogy and frequency, and then measuring the level of learning success. The prosodic patterns would be entirely regular (depending on the compound type) or irregular to different degrees.

We have outlined some important areas for future research on L2 prosody, but undoubtedly there are many more in this yet under-studied area of L2 acquisition.

References


Jaeger, T.F. 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4): 434-446.


**Appendix**

- Kraft faced a class-action lawsuit over its dairy products. The firm settled that case, admitting no wrong doing but agreeing to set up a fund to reimburse unhappy cheese eaters.

- Our beach house was the place from where I strode proudly on summer days into the surrounding neighborhood of graceful houses, to pursue my career as a wage earner.

- “I have visited five shops in search of a packet of sugar but it is not available anywhere,” declared Ahmed, a resident of Township. He said the shopkeepers had stopped sugar selling, fearing government crackdown.

- Mr. Chen, a *Harvard trained* lawyer, was sentenced to life in jail for money laundering and *bribe taking*. Mr Chen’s supporters claim a political vendetta, orchestrated by Mr. Ma. This is hardly in keeping with the clean-guy image of Mr. Ma.
• Thanks to a micro-finance loan, Mary was able to buy a basic handset and a roof mounted antenna to ensure a reliable signal.

• Since 1945, the United Nations and their agencies have been collecting extensive country based data from member countries and generating valuable information on a variety of topics.

• We visited an apple farm located in the lush mountains of western North Carolina. They provided us with samples of their produce and their specialty applesauce.

• During the winter season, most specialty fruit stores sell Blood oranges from Morocco. They are very sweet and make a great breakfast treat.

• People were flocking to the town hall for some noon dancing even though it was raining cats and dogs.

• Wrapping your baby is one of the most effective ways to increase the length of their nap. Give it a try, you may well be surprised on the effect on your baby’s day sleeping.

• We have noticed that city raised cats do not get frightened easily because they are used to hearing police sirens, traffic noises, and voices.

• The library is hosting a number of events to celebrate this year's Children's Book Festival. On Friday there will be Farmyard Tales with an afternoon of stories from the farm. Children should bring their own toy pet for that event.

• Thanks to newly installed village phones, people in remote areas of the country can have access to basic communications.

• In this gallery, you'll find many of the usual stone buildings: for example, castles, churches, and courthouses of sandstone and limestone.
• Johnny was looking for the *metal box* that had many tools in it, including a hammer, a drill, and a battery pack.

• If you don't already do so, consider keeping a headache diary. After a migraine passes, write down when and how your *headache* began and what might have set it off.

• My wife fed our baby daughter a spoonful of liver baby food. Our daughter spit the food half way across the room. My wife picked up our daughter from the *highchair* and gave her a cookie.

• Kids and families pack empty *shoeboxes* with simple items most people take for granted, such as *toothpaste*, toys and school supplies.

• Jane was supposed to feed a *goldfish* down the street for a family when they were out of town.

• Jim Jones was reinstated after the fire commissioner granted an exception to a ban on locker decorations. Jim is a suburban Philadelphia *firefighter* who was suspended over a *flag sticker* on his locker.

• Professor Anderson likes to write on the *blackboard* when explaining mathematical proofs and equations.