### The Cost of Prosodic Markedness as a Rationale for the Placement and Interpretation of Accents

In light of experimental evidence that accent placement cannot be predicted from properties of the discourse context alone (German et al., in press), this paper demonstrates that a mutual representation of the speech situation with regard to the expected utility of outcomes provides a rationale for the speaker's choice of and the hearer's interpretation of accent placement. The cost of using marked forms is shown to interact productively with the risk of miscommunication, giving rise to explicit predictions about speaker and hearer behavior in certain discourse situations. While preserving the coverage of earlier work on accent placement and interpretation (Selkirk 1995, Schwarzschild 1999), I show that a game-theoretic approach has empirical advantages over alternative accounts in several respects.

Most theories of accent placement rely on the assumption that the mapping from discourse contexts to accent pattern is many-to-one. Consequently, they provide no basis for predicting speaker and hearer behavior in cases where the grammar provides optionality with regard to the accent patterns that are available to the speaker. German et al. (in press) showed that the mapping from discourse context to accent pattern is, in fact, also one-to-many in certain contexts. The study used an experimental paradigm to elicit speakers' preferred accent patterns in carefully controlled discourse contexts. In particular, it looked at the interaction between an utterance's potential to signal the discourse status of referents, and independent prosodic constraints on the locations of accents (prosodic well-formedness). The interaction is best illustrated in (1B), which crucially involves a relative clause ending in a sequence of a verb, a NP, and a stranded preposition.

(1)     A: Could I ask you to slice the bread?
        B: First I need a knife to <u>slice the bread with</u>.

The standard intuition, and the prediction made by standard theories of focus projection is that in (1B), the nuclear accent should fall on the preposition (*with*), and the NP (*bread*) should remain unaccented in order to signal that the information expressed by the PP is New and that expressed by the VP (including the NP) is Given. German et al. (in press), however, showed that while speakers unanimously opted to place the accent on the NP when it was New (not shown), they alternated between placing an accent on the NP (70% of trials) and placing one on the preposition (30% of trials) when the NP was Given as in (1B).

These results show rather conclusively that the grammar provides some optionality with regard to a speaker's choice of accent placement. In particular, they show that when an utterance has the general syntactic form of (1B), and occurs in a *narrow focus* context (i.e. the VP is Given), the speaker can choose to accent either the preposition or the NP. For the same sentence in a *broad focus* context (VP is New), placing an accent on the preposition appears to be ruled out. This particular relation between meaning and form is illustrated in (2).

(2)     NP is New  ←————————————→  Accent on NP

        NP is Given  ←————————————→  Accent on P
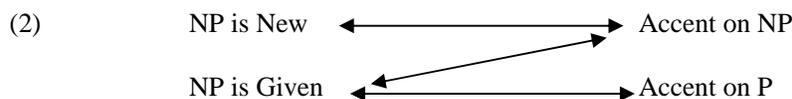
These results are problematic for available theories of accent placement, which make no predictions about how a speaker chooses between forms or how a hearer chooses between interpretations in such a situation. In what follows, the case is made that if two kinds of information are mutually available to the speaker and hearer, namely, a probability distribution over the two types of discourse status the NP can have and a utility function over the possible outcomes, then the behavior of speakers and hearers with regard to accent placement can be predicted based on the equilibria that emerge from the possible combinations of strategies available to them. To see why this is so, first consider the actions that are available to the speaker and the hearer, shown in (3).

(3)     $f_{NP}$: speaker places an accent on the NP
        $f_P$: speaker places an accent on the preposition
        $m_N$: hearer identifies the NP as New
        $m_G$: hearer identifies the NP as Given

If $S_N$ and $S_G$ stand for the two initial situations that a speaker and hearer could be in with regard to the discourse status of the NP ($S_N$=New, $S_G$=Given), then the following strategies are available to them:

(4) | Speaker Strategies | Hearer Strategies
--- | --- | ---

$s_1$: $\quad S_N \longrightarrow f_{NP}$ $\qquad\qquad$ $h_1$: $\quad f_{NP} \longrightarrow m_N$
$\qquad\quad S_G \longrightarrow f_P$ $\qquad\qquad\qquad\qquad\quad f_P \longrightarrow m_G$

$s_2$: $\quad S_N \longrightarrow f_{NP}$ $\qquad\qquad$ $h_2$: $\quad f_{NP} \longrightarrow m_G$
$\qquad\quad S_G \longrightarrow f_{NP}$ $\qquad\qquad\qquad\qquad\quad f_P \longrightarrow m_G$

Following Parikh (2001), it is assumed that the speaker knows prior to the utterance which of $S_N$ and $S_G$ holds. A speaker's strategy, therefore, is a way of choosing an utterance given a situation, and hearer's strategy is a way of choosing an interpretation given an utterance. The shared goal is to make the hearer's choice of actions ($m_N$, $m_G$) coincide with the situation they both start out in ($S_N$ or $S_G$). Importantly, $m_N$, $m_G$ are assumed to share the same propositional content, and differ only with respect to the discourse status associated with the NP.

If this were a *cheap talk* game, where all utterances have the same cost, then speakers and hearers would always select the pair of strategies $<s_1, h_1>$, since that would guarantee successful communication. German et al. showed that this cannot be the case, however. We therefore need some way of representing how the utilities associated with outcomes depend on the strategies that were used to arrive at those outcomes. For simplicity, I assign a utility of +1 to the best possible outcome in which both the propositional meaning and the discourse meaning are communicated successfully. All other outcomes are defined relative to this.

We can represent the cost of a mismatch between the actual discourse status of an NP and the interpretation a hearer chooses as $-\alpha$. Briefly put, this cost derives from the increased risk of needing to initiate a repair strategy in a subsequent game.[1] The cost of the utterance forms, $f_{NP}$ and $f_P$, follows from their relative markedness. In other words, utterances that include accents on highly functional, prosodically non-prominent words such as prepositions are marked relative to forms with accents only on prosodically prominent content words like nouns. It is argued that relative to a less marked form, a more marked form will have some negative utility associated with the higher cost of processing it, which is represented by $-\varepsilon$. Support for this assumption is found in Terken & Nooteboom (1987).

If a speaker and hearer have mutual knowledge of the probability distribution P over $S_N$ and $S_G$, where $P(S_N) + P(S_G) = 1$, then the expected utility of each possible outcome is given in (5)-(8).

(5) $\qquad <s_1, h_1>$: $\quad EU = P(S_N)*(1) + P(S_G)*(1-\varepsilon) \quad = 1 - \varepsilon*P(S_G)$

(6) $\qquad <s_1, h_2>$: $\quad EU = P(S_N)*(1-\alpha) + P(S_G)*(1-\varepsilon) = 1 + \alpha*P(S_G) - \alpha - \varepsilon*P(S_G)$

(7) $\qquad <s_2, h_1>$: $\quad EU = P(S_N)*(1) + P(S_G)*(1-\alpha) \quad = 1 - \alpha*P(S_G)$

(8) $\qquad <s_2, h_2>$: $\quad EU = P(S_N)*(1-\alpha) + P(S_G)*(1) \quad = 1 + \alpha*P(S_G) - \alpha$

Given that $\alpha, \varepsilon > 0$, it can be shown that the following inequalities hold for the above expressions:

(i) $\qquad s_1h_1 > s_1h_2$

(ii) $\qquad s_2h_2 > s_1h_2$, $\qquad$ for $P(S_G)>0$; otherwise $s_2h_2 = s_1h_2$

(iii) $\qquad s_1h_1 > s_1h_2$, $\qquad$ for $P(S_G)>.5$; and $s_1h_1 < s_1h_2$, for $P(S_G)<.5$

(iv) $\qquad s_1h_1 > s_2h_1$, $\qquad$ for $\alpha > \varepsilon$; and $s_1h_1 < s_2h_1$ for $\alpha < \varepsilon$

(v) $\qquad s_2h_2 > s_1h_1$, $\qquad$ for $P(S_G)> (\alpha \div (\alpha + \varepsilon))$; and $s_2h_2 < s1h1$, for $P(S_G)< (\alpha \div (\alpha + \varepsilon))$

The effect of (i) and (iv) together is that as long as the cost of miscommunicating discourse status ($\alpha$) is greater than the cost of using a marked form ($\varepsilon$), then the combination of strategies in which the speaker uses a marked form for Given NPs and the hearer is sensitive to the utterance being used is guaranteed to be a Nash equilibrium. Intuitively, this is the right result. If it cost more to use marked forms than to accept the

---

[1]The assumption is that an increase in risk of future cost can be represented as a negative utility in the game under consideration. Other ways of construing $\alpha$ are discussed in depth in the full version of the paper.

consequences of miscommunication, then unmarked forms would be used exclusively. In other words, speakers and hearers would always choose either $s_2h_1$ or $s_2h_2$. This is, in fact, not consistent with the results of the German et al. study, since speakers were observed to use marked forms in $S_G$ approximately 30% of the time. We conclude, then, that $\alpha > \varepsilon$.

|  | $h_1$ | $h_2$ |
|---|---|---|
| $s_1$ | $1-\varepsilon*P(S_G)$ | $1+\alpha*P(S_G)-\alpha-\varepsilon*P(S_G)$ |
| $s_2$ | $1-\alpha*P(S_G)$ | $1+\alpha*P(S_G)-\alpha$ |

**Figure 1 - Utility matrix for accent placement**

Depending on the value of $P(S_G)$, there may be one Nash equilibrium in addition to $s_1h_1$. As long as $P(S_G)>0$, then $s_2h_2$ dominates $s_1h_2$, though it only constitutes a Nash equilibrium if it also dominates $s_2h_1$, which is the case whenever $P(S_G)>.5$. Thus, $s_1h_1$ alone is an equilibrium whenever $P(S_G)<.5$, whereas $s_1h_1$ and $s_2h_2$ are both equilibria whenever $P(S_G)>.5$. The concept of Pareto dominance predicts which combination of strategies a speaker and hearer prefer when $P(S_G)>.5$. (v) tells us that $s_2h_2$ is preferable to $s_1h_1$ whenever $P(S_G)> (\alpha\div(\alpha+\varepsilon))$. Notice, however, that this only matters if $\alpha>\varepsilon$, since otherwise, $s_1h_1$ is not a Nash equilibrium. In fact, as long as $\alpha>\varepsilon$ holds, then also $(\alpha\div(\alpha+\varepsilon)) > .5$. This means that whenever $s_1h_1$ is an equilibrium, then $P(S_G)$ has to be *at least* greater than .5 in order for $s_2h_2$ to Pareto-dominate $s_1h_1$.

In summary, the above results make following general predictions regarding the behavior of speakers in a context where an NP is Given:

(a) Speakers will use a marked accent pattern if the mutual probability of NP being Given is relatively low, and an unmarked pattern if that probability is high.

(b) For a constant cost of using a marked form, speakers will use a marked accent pattern if the cost of miscommunication is very large, and an unmarked pattern if that cost is small and the mutual probability of NP being Given is greater than half.

(c) For a constant cost of miscommunication, speakers will use a marked accent pattern if the cost of doing so is very small, and an unmarked pattern if that cost is very large and the mutual probability of NP being Given is greater than half.

Similar predictions can be made for hearers' interpretation strategies. Because they can be stated in qualitative terms, these predictions can be tested with some degree of empirical precision. After all, German et al. showed that speakers use marked forms variably in contexts like (1), suggesting that $(\alpha\div(\alpha+\varepsilon))$ is neither too high nor too low for speakers to exhibit sensitivity to its value. As an example, since the recency of an NP's last coreferring expression is predicted to correlate positively with $P(S_G)$, it may be possible to test (a) in an experimental setting by manipulating this factor. (b) is potentially testable as well by manipulating the extrinsic consequences that speakers and hearers expect to face when discourse status is miscommunicated (i.e., the value of $\alpha$). In a two-player game setting, for instance, the speaker and hearer could be given a task in which discourse status is either more or less necessary for completing that task. Finally, (c) is testable in a setting where speakers and hearers engaged in a game task are limited in the utterance forms that are available to them, and each form is associated with a different extrinsic cost (i.e., a different $\varepsilon$). These experimental paradigms are currently under development and will be presented in more explicit detail in the final paper.

This analysis relies on relatively few unknown variables (i.e. $\alpha$, $\varepsilon$, and P), and in that sense, makes explicit and testable predictions for the placement and interpretation of accents. It has potential, then, not only as a model for accent placement, but as a way of uncovering and quantifying the utility functions that lie at the foundation of all situated applications of game-theory in linguistics. The methodology being developed for this analysis promises to extend readily to other analyses (e.g., blocking effects (Dekker & van Rooy 2000)) that also involve an interaction between meaning-form alignment and markedness.

**References**

Dekker, P. and van Rooy, R. 2000. Bi-directional Optimality Theory: An Application of Game Theory. *Journal of Semantics* 17: 217-242.

German, J., Pierrehumbert, J. and Kaufmann, S. In press. Evidence for phonological constraints on accent placement. *Language,* 82(1).

Parikh, P. 2001. *The Use of Language*. CSLI Publications, Stanford, CA.

Schwarzschild, R. 1999. GIVENness, AVOIDF and other constraints on the placement of accent. *Natural Language Semantics* 7.141-77.

Selkirk, E. O. 1995. Sentence prosody: intonation, stress, and phrasing. *The handbook of phonological theory*, ed. by John A. Goldsmith, 550-69. Cambridge MA: Blackwell.

Terken, J. and Nooteboom, S.G. 1987. Opposite effects of accentuation and deaccentuation on verification latencies for Given and New information. *Language and Cognitive Processes 2*, 145-163.