'A Statistical Analysis of Acceptability Judgments'

1. **Introduction:** A quite common practice among generative syntacticians (in a broad sense) is to ask their informants to compare two sentences constituting a minimal pair. When one sentence is reported to be noticeably less acceptable than the other, it is regarded (more often than not) as ungrammatical. This practice, however, is quite dangerous since there are so many extra-syntactic/grammatical factors that can lower acceptability of linguistic expressions. As is well-known, quest for grammaticality judgments becomes legitimate only when researchers succeed in distilling grammaticality effects from such heterogeneous acceptability judgments provided by language users — an absolutely necessary but extremely difficult process often neglected in the 'idealization' approach of generative tradition. Moreover, when researchers' opinions differ as to the grammaticality of some particular sentences or sentence types, it is often treated, rather casually, either as mere idiolectal variation, or it is simply left irreconcilable. Our goals of this work are stated in (1):

(1) By conducting experiments on the acceptability judgments made by a substantial number of native speakers, and statistically analyzing their results, we will attempt:

   a. To identify multiple factors and their interactions contributing to acceptability judgments.
   b. To demonstrate that statistical analyses of acceptability judgments can resolve some of the seemingly irreconcilable issues of grammaticality judgments, especially concerning the discrimination between ungrammaticality and unacceptability.

What we explore in this work, in other words, is not any new syntactic analysis but is what we believe is an effective method of ascertaining the nature of speakers' acceptability judgments when they are controversial or questionable.

2. **A model case in Japanese:** Japanese has a syntactic causative construction which involves a morphologically complex causative predicate of the form V-**sase** 'V-CAUSE' as in (2), some of whose major characteristics are summarized in (3):

(2) Taro-ga kodomo-*ni/-i o ik-ase-ta.  
    Taro-NOM child-DAT/ACC go-CAUSE-PAST

(3) a. At surface, the sentence is syntactically **monoclausal:**  
    \[ [\text{IP Causer} \ \text{Causee} \ \text{V-\text{**sase**}}] \]

   b. Its semantic interpretation suggests that a **biclausal** structure must be involved at some point of syntactic derivation:

   \[ (i) \ [\text{IP Causer} \ \text{Causee}_1 \ [\text{XP [e]}_1 \ \text{V \ **sase**}] \text{ and/or} \]
   \[ (ii) \ [\text{IP Causer} \ [\text{XP Causee}_1 \ \text{V \ **sase**}] \text{ and/or} \]

   c. Causees can be marked either with -*ni* 'DAT' (= **ni-causative**) or -*o* 'ACC' (= **o-causative**).

One point of controversy has been whether and how the syntactic positions of *ni*-causee and *o*-causee differ in the complex syntactic structure in (3) — some analyze the *ni*-causee as a **matrix** object as in (3b-i) ((Toniike (1978)), the *o*-causee as a **matrix** object ((Kuroda (1965)), and both causees as a **subordinate** subject as in (3b-ii) ((Kitagawa (1986))). Accordingly, different outcomes are predicted concerning the way verb-modifying adjuncts are interpretable when they are located to the right or left of *ni- or o-causee in (3b-i) or (3b-ii).

Included in *The Handbook of Japanese Linguistics* (1999: Blackwell) is a highly controversial analysis proposed by Miyagawa (1999), which deals with this issue. It introduces a morpheme for a causativized predicate (V-**CAUSE**) en route to PF a la Distributed Morphology, and both *ni- and o-causatives are claimed to maintain a **biclausal** construction throughout the syntactic derivation but in an asymmetrical fashion as in (4a-b) — henceforth, this analysis will be referred to as the **DM-analysis**:

(4) a. Causer-ga Causee-*ni \ [\text{IP [e]}_1 \ \text{V \ **sase**}] \text{ and/or} \]
    \[ -\text{NOM} \ -\text{DAT} \text{ CAUS} \]
   b. Causer-ga Causee-*o \ [\text{IP Causee-\text{o}} \ \text{V \ **sase**}] \text{ and/or} \]
    \[ -\text{NOM} \ -\text{ACC} \text{ CAUS} \]

This analysis predicts that an adjunct that modifies the subordinate verb (V-**A**\text{-}**j**\text{et}) fails to appear to the **left** of the *ni*-causee at surface as illustrated in (5a) below, and inevitably also that an adjunct modifying the causative predicate in the matrix (C-**A**\text{-}**j**\text{et}) fails to appear to the **right** of the *o*-causee as in (5b).
(5) a. Causer-ga \( V_{-}Ajet \) Causee,\( -ni \) [IP [e] \( V \) ]-sase \\
\| \[ ] \| \\
b. Causer-ga \[ [IP Causee-o C_{-}Ajet ] V \] -sase \\
\| \[ ] \| 

The contrast in (6) is reported as support for this analysis. ((6a-b) are from Miyagawa (1999: 249) including judgments and translations. We disregard other aspects of the DM-analysis that are irrelevant to our discussion.)

(6) a. 0K \( Taro-go \) [IP \( V_{-}Ajet \) kooen-e ] kodo\( m-o \) [\( V \) ik ] ]-ase-ta. 'Taro made (his) child go to the park.'
\( Taro-NOM \) \( park-to \) \( child-ACC \) \( go-CAUSE-PAST \)

b. ?? \( Taro-go \) [IP \( V_{-}Ajet \) kooen-e ] kodo\( m-o \) \( ni \) [IP [e] \( V \)] \( ik \) ]-ase-ta. 'Taro let (his) child go to the park.'
\( Taro-NOM \) \( park-to \) \( child-DAT \) \( go-CAUSE-PAST \)

3. Experiments: We conducted experiments described in (7) below to examine this prediction. Our main goal here is to demonstrate the usefulness of statistical analyses in discriminating speakers' recognition of ungrammaticality from that of lowered acceptability induced by extra-grammatical factors, rather than refuting the DM-analysis in particular.

(7) a. 58 subjects, all non-linguists, native speakers of Japanese, ages 18-20.

b. Pre-test: 24 sentences to test if the subjects actually distinguish \( V_{-}Ajet \) from \( C_{-}Ajet \. 51 \) reliable subjects were identified. (7 persons who failed to clearly discriminate \( V_{-}Ajet \) from \( C_{-}Ajet \) were excluded.)

c. Main experiment: Causative sentences containing an adjunct with 8 distinctive combinations of 3 aspects of causative sentences — [1] Adjunct type (\( V_{-}Ajet \) vs. \( C_{-}Ajet \), [2] Particle type (\( ni \) vs. \( o \) on causee), and [3] Location of \( Ajet \) (left vs. right of causee), 6 sentences each, totaling 48.

d. The subjects rated 48 experimental and 24 filler sentences by selecting one of: (a) clearly legitimate, (b) basically legitimate, (c) perhaps illegitimate, (d) clearly illegitimate, with some general instructions and examples clarifying the notion of "legitimate sentences" provided in advance. (a)–(d) in the subjects' answers were converted into scores of 4–1 points, respectively.

In all experimental sentences, the sequence of an adjunct and a causee (in either order) was placed in the background portion of a pseudo-cleft construction as in (8) below in order to discourage the focalization of the adjunct, which Miyagawa considers to be the reason why (6b) may not be regarded as fully unacceptable.

(8) [Background … \( hiraoyogi-de \) \( sensyutati-ni \) oyog-ase-ta-no-wa ] [Focus \( ano-kooti-desu \) ]
\( breast.stroke-in \) \( swimmers-DAT \) \( swim-CAUSE-PAST-HE.WHO-TOP \) \( that.coach-COPULA \)

'It is that coach who made the swimmers do the breaststroke.'

We obtained the findings and conclusions summarized in Sections 4-7 below.

4. Factor analyses: In order to examine in what way sentences and speakers are divided into subgroups, we conducted Factor Analysis which compared the patterns of acceptability responses concerning distinct sentence types by each subject, using (9).

(9) Extraction Method: Unweighted Mean Square, Rotation: Varimax, 3 Factors with Eigenvalues higher than 1 were extracted, \( R^2 = .766 \) (Factor 3 identified the 7 subjects mentioned in (7b) above.)

(10) below illustrates all sentences plotted by Factor 1 and Factor 2 scores — because the number of sentences is larger than that of subjects in our experiment, we put the subjects in columns and the sentences in rows in the data set for our Factor Analysis.
[1] Factor analysis of sentences: Factor 1 clearly distinguishes grammatical fillers & grammatical pre-test sentences (= (i) in (10)) from ungrammatical fillers & ungrammatical pre-test sentences (= (ii)). Crucially, all experimental causatives are grouped with (i). (Mean acceptability: (i) = 3.60, (ii) = 1.53 (possible range of acceptability = 1.00 ~ 4.00)) That is, Factor 1 discriminates grammatical sentences (ii) including all causative sentences under all distinct experimental conditions described in (7c) from ungrammatical sentences (iii).

Factor 2 distinguishes, in a less clear fashion, all grammatical fillers & grammatical pre-test sentences, o-causatives and 20% of ni-causatives (= (iii) in (10)) from 75% of ni-causatives (= (iv)) within the group (i) above. The ni-causatives in (iii) are highly acceptable while those in (iv) are less acceptable. (Mean acceptability: (iii) = 3.78, (iv) = 3.15) Factor 2, in other words, distinguishes grammatical and highly acceptable sentences (iii) from grammatically but somewhat less acceptable sentences (iv).

[2] Factor analysis of speakers: Factor 1 is correlated with each subject's degree of discriminating the grammatical sentences (i) from ungrammatical sentences (ii) above. The higher the Factor 1 load of each subject is, the more clearly she accepts sentences of group (i). (Pearson's r (Factor 1 * Mean acceptability of sentences in (i) by all subjects) = .800)

Factor 2 is correlated with the degree the subjects were affected by whatever awkwardness grammatical sentences (i) may involve. The higher the Factor 2 load is, the lower the acceptability of some ni-causatives ((iv)) becomes. (Pearson's r (Factor 2 * Mean acceptability of (iv) (= 75% of ni-causative) by all subjects) = -.634) This phenomenon is well in accordance with the oft-reported observation that ni-causative is pragmatically more restricted than o-causative. Apparently, Miyagawa's ni-causative example (6b) and its adjusted version (8) belong to (iv).

5. Subgroups of Subjects: Factor Analysis identified two speaker groups. Both Group 1 and Group 2 clearly distinguish grammatical sentences from ungrammatical sentences, but Group 2 is more affected by the (arguably extra-grammatical) awkwardness in their acceptability judgment than Group 1, and deems grammatical but somewhat awkward sentences ((iv)) to be lower in acceptability. (Mean acceptability of (iv) (= 75% of ni-causative): Group 1 (32 subjects) = 3.43, Group 2 (19 subjects) = 2.69)

6. Interaction of particle types & Aject locations — 2-way ANOVA:

[1] Particle effect: Acceptability is lower with ni-causee than o-causee irrespective of the position of Aject in relation to the causee. (Main effect of particles: V-Aject — mean acceptability of o = 3.88, ni = 3.32, F(1, 305) = 218.00, p < .001 / C-Aject — mean acceptability of o = 3.81, ni = 3.22, F(1, 305) = 161.3, p < .001)

[2] Particle*Location Interaction: When the adjunct is located to the left of ni-causee, acceptability tends to be lower. (Interaction (particle*location): V-Aject — F(1, 305) = 66.81, p <.001; mean acceptability of "V-Aject to the right of ni" = 3.57 / "V-Aject to the left of ni" = 3.07) cf. The lower acceptability of (5b)/(8).

[3] No Aject effect: [1]-[2] above are observed for not only V-Aject but also C-Aject. (Interaction (particle*location): C-Aject — F(1, 305) = 28.47, p < .001; mean acceptability of "C-Aject to the right of ni" = 3.33 / "C-Aject to the left of ni" = 3.10)

[4] High acceptability: All experimental conditions yield mean acceptability higher than 3. (Range of mean acceptability of sentences under all experimental conditions: 3.07 (V-Aject to the left of ni) ~ 3.89 (V-Aject to the right of o))

7. Conclusion on the DM-analysis: Thus, no speaker provided the grammaticality judgments predicted by the DM-analysis, and the two claims of the DM-analysis on the adjunct-particle-location interaction (5a-b) fail to be supported. The informants whose judgments were used to support the DM-analysis are likely to belong to Group 2, and the lower acceptability they reported, presumably, was misinterpreted as ungrammaticality. Further experiments and analyses along this line are being planned for a few more empirical phenomena that involve similarly controversial judgments.