

Frequency effects of case-markers on focushood —evidence from dialectal difference in Japanese—

Category: cognitive/functional syntax/semantics

1 Introduction: target phenomenon

This work examines a corpus consisting of the two major dialects of Japanese (the Eastern dialect or *kantôben* and the Western one, *kansaiiben*) from the viewpoint of information structure and presents statistical evidence that as regards the case-markers *ga* and *o*, the frequency of their overt occurrences is significantly different between these dialects *if conditioned on the argument structure*. The main finding is that there is a difference in overt-marking frequency not only between various argument structures within a dialect but also within an argument structure between the dialects. In view of this result, coupled with the observation on the perception of focus (described immediately below), we develop the account that each dialect has its own overt-marking tendency in a way relative to argument structures, and the perception of focus articulation is influenced by this tendency (last section in this abstract).

It has been well-attested that the case-markers under investigation, particularly *ga*, carries a strong narrow focus effect for a subset of predicates. For example with statives, including copula (*da* and its variants) and adjectives, *ga* can only be felicitous in a narrow-focus context (as in b-1 below) while the wide (or sentence) focus reading is possible for some other predicates (as in b-2).

- (1) (each of the b sentences should follow a)
- a. Kinô atarashii shain-ga haitta.
yesterday new staff-NOM joined
'A new member of staff joined yesterday'
- b-1. ?? Sono shain-ga Amerika-jin da. (narrow focus only)
that staff-NOM American COP
'It is that member of staff that is American'
- b-2. Sono shain-ga kyô chikoku-shite kita. (both narrow and wide focus available)
that staff today arriving late came
'That member of staff was late for work today'

A curious point is that the context cannot override the narrow-focus effect in (1b-1). This might lead to the view that *ga* itself inherently has this effect, but such a possibility is contradicted by (1b-2). The effect, rather, seems relative to the argument structure types, as observed in [2].

However, this is the observation in *kantôben*, the dialect mostly discussed in the literature, and the situation is different for *kansaiiben*. For a sentence like (b-2), many *kansaiiben* speakers report that overt case-marking feels redundant or even inappropriate. Consider the following variants of (1b-2), one in *kantôben* and the other in *kansaiiben*:

- (2) kantô. Kyô kinô haitta shain- $\{ga/\phi\}$ chikoku-shite kitandayone.
kansai. Kyô kinô haitta shain- $\{\phi/ga\}$ chikoku-shite kiyotten.
(‘The member of staff who joined yesterday was late for work today’)

The contrast between the dialects seems to amount to this: while in the narrow focus context both sets of speakers seem to prefer to use an overt *ga* (say, to answer the question *Who arrived late?*), in the wide-focus (e.g. discourse-initial) context, the overt *ga* tends to be preferred over zero-marking in *kantôben*, but the preference is reversed in *kansaiiben*. The fact that *kansaiiben* speakers at times find an overt marker ‘inappropriate’ suggests that its presence forces the narrow-focus reading more often in this dialect, while there is no such effect in *kantôben*, even in exactly the same context. All this points to the conventionalised nature of the effect of case-marking on focushood, relative to argument structure. In what follows we present corpus evidence to support this view and offer a possible account.

2 Data and findings

To identify what variables are closely correlated to the presence/absence of our case-markers, we have investigated our self-compiled corpus consisting of two subparts, *kansaiiben* and *kantôben* (approx. 2,500 sentences each), all taken from the performances of *manzai*, a genre of stand-up comedy in which a pair of performers engage in a comical dialogue, a type of materials we consider conducive to information-structural

motivations and case-marker ellipsis. As we focus on two case-markers, the target data was their subsets: we looked at subject/object arguments which are either *ga*-marked, *o*-marked or zero-marked.

We first conducted two simple correlation tests (χ^2) to estimate the effects on case-marking of, the dialect difference on one hand, and the argument structures on the other. Somewhat surprisingly, the first test (correlation of the overall counts of overt case-marking and the two dialects) did not produce a great contrast either for *ga* ($p = .0605$) or for *o* ($p = .0726$). The naive observation that *kansaiben* speakers drop a case-marker more often was not confirmed. On the other hand, the second comparison, i.e. between the marker occurrences and the argument structure types (values: unaccusative, unergative and transitive), produced a significant result in *kansaiben* ($p = .0378$), and a slightly less significant one in *kantôben* ($p = .0502$).

To ensure more general results, so as to encompass multiple predictor variables (aka. ‘factors’ or ‘independent variables’), which in our case are, at the very least, *galo* and argument structures, and further, to assess the difference between the two dialects, we then invoked the statistical technique of *logistic regression*, the advantage of which is that one can combine potentially inter-dependent predictor variables and computes the result in continuous terms: how probable overt case-marking is, in our case. We then compared the probabilities between the two dialects.

The table below shows the probability of case-marker appearance (as opposed to zero-marking) for the relevant combinations. We also show the χ^2 significance levels for the differences between dialects and between combinations, as well as the overall data counts (i.e. the number of subject/object NPs with overt-marking or zero-marking, *not* the overtly-marked NP counts), as they have bearing on the statistics.

	<i>kantôben</i>	<i>kansaiben</i>	sig. bet. dialects	overall datapoints (<i>kantô</i> / <i>kansai</i>)
<i>ga</i> /unergative	53.25%	52.36%	n/s	501/419
<i>ga</i> /unaccusative	55.63%	46.85%	***	359/388
<i>ga</i> /transitive	74.43%	71.14%	**	1154/1209
<i>o</i> /transitive	50.41%	48.57%	n/s	ditto
sig. bet. combinations	**	**		

As can be seen, the result is ‘mixed’, but this is what we expected: there appears to be a difference in terms of correlation for some argument structure / case combinations but not for some others.

3 An account and some implications

An account that could capture the above data is that it is an *interaction* of the argument structure of a predicate *and* the (in)frequency of the case-marker occurrences that determines the actual perception of focushood on the argument NPs. Neither appears sufficient on its own. The argument structure would not account for the difference in the perceived focushood in different dialects, and with the frequency alone, the fact that our investigation did not produce significant difference for either case-marker would be left amiss. However it would make sense to say the presence/absence of a case-marker is the surface indicator of which argument(s), if any, is normally or *by default* focused for a type of predicate, in the sense that if the focus falls on the *predictable* position, the speaker simply follows the convention. In the case of our dialectal contrast for unaccusative, while the ‘default’ reading is wide focus in both dialects, the convention is zero-marking with subject in *kansaiben* and overt-marking in *kantôben*. We then say a deviation from this convention is perceived as the deviation from the default reading, and hence the narrow focus effect ensues in *kansaiben*. This also predicts for *kantôben*, in our view correctly, that dropping a case-marker could have ‘de-focusing’ effect, and also that, some other means such as prosody is required to explicitly indicate narrow focus.

This is essentially an information-theoretic (or entropy-based) account, akin to [1], according to which one does not bother to do anything unusual unless unpredictable. This type of account can be contrasted with ones based on discourse contexts (e.g. [4]) or inherent properties (such as animacy) of NPs [3], although we certainly do not exclude the influences from these factors. We say, as it were, a major determinant for case-marking is the collective memory incalculated in a speech community, rather than the spot decision made for an individual sentence in a particular context with particular types of NPs. Such an account could be extended to a similar difference across languages, e.g. the case-marking convention difference between Japanese and Korean, where the (de-)focus effects of the subject case-marker are markedly different.

References

- [1] F. Jaeger. Redundancy and reduction: Speakers manage information density. *Cognitive Psychology*, 61(1):23–62, 2010.
- [2] T. Kageyama. *Bunpô to Go-keisei (Grammar and word formation)*. Hitsuji Shobo, 1993.
- [3] H. Lee. Case ellipsis at the grammar/pragmatics interface: A formal analysis from a typological perspective. *Journal of Pragmatics*, 39(9):1465–1481, 2007.
- [4] T. Ono, S. Thompson, and R. Suzuki. Pragmatic nature of the so-called subject marker *ga*. *Discourse Studies*, 2000.