

Markednesses

Henk Zeevat*

1 Overview

Two developments make the concept of markedness important. One is the availability of frequency statistics and the rise of probabilistic methods in computational linguistics, but now also increasingly into linguistic proper. Probability naturally gives rise to its own notions of markedness. The second development is optimality theory, a formal theory of markedness. This paper tries to analyse notions of markedness. In addition, it tries to critically assess the assumption of markedness harmony: if something is marked in way A, then it is also marked in way B and the historical tendency formulation of this: in history, such harmony increases. I will try to show that markedness concepts are only rarely in harmony and that history can also destroy harmony. My aim was to reduce confusion, especially my own, and to understand better the relation between probabilities and optimality theory. I am not sure I have succeeded.

2 Markednesses

Semantically Marked A meaning can be marked with respect to another in several ways. *She-donkey* has a meaning which is marked with respect to the meaning of donkey, because it contains the meaning of donkey and then adds an extra feature: female. If we say: *That is not a donkey, it is a she-donkey donkey* itself acquires the marked meaning *he-donkey* due to the contrast relation into which it enters with *she-donkey*. This is essentially what Jakobson [7], [8] had in mind:

If an expression can be interpreted in several ways, an interpretation is marked if another possible interpretation is properly contained in it.

A second notion of semantic markedness can be derived from optimality theoretic constraints such as **Don't Overlook Anaphoric Possibilities** [6], **Do**

* I wish to thank Paul Smolensky, Rusiko Asatiani, Jens Allwood, Oren Schwarz and Hanjung Lee for helpful critical discussion of this material.

not Accommodate ([2]) or ***Invent** [11]. The principles all enforce a preference for old interpretations if they are possible, with old defined by the referential hierarchy ([5]).

If an expression can be interpreted in a number of ways, one interpretation is marked if it is less activated than another possible one.

The next notion that I find reasonable derives from probabilistic parsing. There the general principle is to hunt for the most probable parse/interpretation.

An interpretation of an expression is marked if it has more probable interpretations.

Sometimes it is also possible to use a typological criterion of semantic markedness for functional words. For example, subjects of the same clause are typologically unmarked antecedents for reflexives, whereas objects of the same clause or subjects of higher clauses are possible in some but not all languages.

An interpretation of a functional category is typologically marked with respect to another interpretation if it is only possible in languages when the second interpretation is also available.

The first and second notion do not seem to conflict and can be combined into “semantic markedness”. There can be conflicts between semantic markedness and the typological and probabilistic notions however.

Formal Markedness Formal markedness can be defined on a number of levels. In *phonology*, we can distinguish phonemes by inherent markedness in terms of articulation effort and similar for syllables and words. Markedness in this sense directly correlates with articulatory effort though it is maybe a bit unclear how the different composing factors should be weighed. In principle, the following is a sequence of increasingly marked syllables.

$ta < it < irt < itr < trg < rtg$

But the number of syllables (or the number of phonemes) can also be taken as generating phonological markedness, so that a partial order is the ultimate result. I will not attempt a definition. In *morphology*, the criterion is very simple. We compare words and the marked word is the unmarked one with extra morphs. An expression $X + morph$ is more marked than X . In *syntax*, markedness is not as clear. It is customary to talk about marked and unmarked word order. This can be understood in terms of a canonical word order or in terms of prominence. A language may have a canonical word order and marked orders are then violations of this canonicity. It seems possible to define canonical word order in terms of probabilities for a particular language. Alternatively, one can use the general principle that prominent things come first, with prominence defined as in [1] (and allowing for language variation in the relative importance of prominence dimensions) .

There is also the same probabilistic notion as we had for semantics. For small bits of semantics, the more probable forms can be distinguished from the less probable ones. It is not a practice in computational logic to look at these proba-

bilities (probabilities are not important for generation systems) but the principle is the same. One would expect the most probable form to be the best form for the meaning.

A constituent is *distributionally marked* iff it has a more limited or deviant distribution with respect to the normal members of its syntactic class. (A similar notion can be defended for morphology: the range of nonsyncretic morphemes a word allows.) This notion has been defended recently especially by [4] and [3]. It connects to one dimension of grammaticalisation. For example, a temporal adverb has a normal adverbial distribution, a temporal particle can have more restricted distribution, which becomes still more restricted or deviant when it becomes a clitic, while as a temporal morpheme it has the most limited distribution of all. I will not include distributional markedness in my discussion and instead make the quite problematic assumption that there is a uniform notion of formal markedness.

Typological Markedness and Optimality Theory Typological markedness has a relatively clear interpretation in terms of conditional universals of language: for all L , if A holds of a language L , B also holds of it. There is a direct relation with the idea of a ranked constraint in optimality theory.

The following three statements seem reasonably well-attested conditional universals.

If a language has passive forms, it has active forms

If a language has zero pronouns, it has full pronouns.

If a language allows topical antecedents for a full pronoun, it allows non-topical antecedents for a full pronoun.

In optimality theory, these can be uniformly implemented by the assumption of universal constraints prohibiting the forms in the condition: ***PASSIVE**, ***ZERO&PRO**, ***TOPIC&FULL** or constraints implying these subcases. In languages with passive, ***PASSIVE** is overridden by other constraints, etc. Smolensky (p.c. and in his forthcoming book) holds the following view of OT constraints:

Every OT constraint expresses typological unmarkedness.

The principle is a direct consequence of the free rankability and the universality of constraints which predicts that the pattern enforced by the constraint occurs in all languages, whenever other constraints do not override it. This still allows for the logical and practical possibility that a constraint can be fully obliterated in some languages.

Obliteration makes the claim look weaker than it is. Since there is a universal set of constraints, we can in principle prove whether obliteration is possible for individual constraints and determine the rankings under which it occurs. A counterexample to the idea that a constraint expresses the typologically unmarked case needs a proof that the constraint cannot be obliterated. An alternative is to redefine typological markedness by pseudo-conditional universals:

Where the constraint cannot be obliterated, every language which has alternatives to the pattern promoted by the constraint also has the pattern promoted by the constraint.

An example of a constraint that can be obliterated is **STAY**, e.g. if there is in a language constraint stronger than **STAY** that forces subjects to move. Or suppose there is a universal constraint that requires (non-contrastive) topics to come first: **TOPIC** < **-TOPIC**. In English, where word order has been hijacked by theta-marking, such a constraint would be invisible. The stronger constraints enforce other patterns, but these patterns are marked with respect to the pattern described by the constraint.

Optimality theory should not only explain constraints by typological markedness, it should also be able to give explanations for conditional universals, when there is not a particular constraint corresponding to it. It is important to distinguish typological markedness from the category of markedness constraints who promote forms or interpretations that are formally unmarked, i.e. which are short, have no morphological marker, are in canonical order etc. But nonetheless any constraint —also markedness constraints— is related to typological unmarkedness.

3 Harmony

Semantic, formal and typological markedness perhaps can be articulated as I did above. But are they in harmony? This boils down to 6 harmony hypotheses, if we distinguish formal, semantic and typological markedness and do not worry about possible disharmonies within those categories.

1. formally unmarked coincides with typologically unmarked
2. semantically unmarked meanings and formally unmarked forms associate
3. semantically unmarked meanings and typologically unmarked forms associate

One can consider:

4. semantically unmarked meanings are typologically unmarked meanings.
- If the probabilistic dimension is added, there are 3 more harmony hypotheses.
5. typologically unmarked forms are more frequent than typologically marked forms
 6. formally unmarked forms are more frequent than formally marked forms
 7. semantically unmarked interpretations are more frequent than semantically marked interpretations

The interesting fact about these statements is that most of them are false. It is not even clear that they are true as tendencies.

1. *formally unmarked coincides with typologically unmarked*

The absence of the root form in Latin for nouns in several declinations shows that there cannot be a conditional universal that the morphologically unmarked

form always occurs when morphologically marked forms do.

In these declinations Latin always expresses gender and case and the formally unmarked form never occurs. The reason seems obliteration forcing expression of gender and case on roots like *puel* or *ekw*, so there is still a chance for our weaker version of typological markedness.

In the other direction, the zero pronouns give a problem (clitic pronouns or inflection-based pronouns give the same problem). The shorter forms are formally less marked than full pronouns, since they lack morphology and are shorter. Yet, the full pronouns are the typologically unmarked case ([3]).

Bresnan argues that the shorter forms are formally marked —as the expression of a meaning— because they violate form-meaning iconicity (with zero forms or inflection there is component of the meaning for which there is no constituent in the sentence) or canonical word order (clitics go to strange places). Perhaps this is the way¹ to argue that zero pronouns (clitics, inflection) are formally marked, but it would seem that one needs to have a way for weighing syntactic markedness against morphological and phonological markedness.

2. semantically unmarked and formally unmarked forms associate

It has been argued that the simple form “kill” expresses a normal meaning (killing in the normal way) whereas the syntactically and morphologically marked, longer (and infrequent) form “cause to die” expresses unusual killings [2], [10]. Similarly, it is the unmarked form “donkey” which expresses the unmarked meaning and not the complex form “she-donkey”. But it is not always equally clear. English obligatorily marks progressive aspect on finite verbs, but as an opposition, so that the simple unmarked form expresses non-progressive and has a meaning that is as marked as the progressive form. Full oppositions expressed by a marked and an unmarked form are frequent like present versus past in English and other languages or perfect versus imperfect and lead to a situation where the unmarked form has a meaning that is as marked as the meaning of the marked form.

Pronouns give a counterexample in the other direction since in Dutch, English, Italian and Spanish pronouns are special by having case and presumably should count as morphologically more complex in the accusative than a caseless bare NP like “flour” or “oil”.

3. semantically unmarked meanings and typologically unmarked forms associate

The facts discussed by Bresnan form a puzzle for this harmony principle. It seems a correct observation that full pronouns are the typologically unmarked case with respect to weak, clitic, zero and morphologically incorporated pronouns. But it is those marked forms that refer to topical antecedents and those are the unmarked meanings, if we assume the definiteness hierarchy.

Full pronouns are typologically unmarked but receive the semantically marked

1. Distributional markedness is another way.

non-topical meanings in languages that have alternatives.

4. the semantically unmarked meanings are the typologically unmarked meanings.

Topical interpretations for full pronouns are typologically marked: they only occur if the pronoun can take non-topical antecedents as well. Topical interpretations are semantically less marked than non-topic interpretations, yet they get the typologically marked reduced forms as their method of expression.

5. typologically unmarked forms are more frequent than typologically marked forms

I have not counted but zero subject pronouns in Italian or Spanish seem far more frequent than the rare subject pronouns. And quite the same holds for clitic object pronouns in French: they are the standard case, not the nearly impossible full forms.

6. formally unmarked forms are more frequent than formally marked forms

This is plainly false. Non case-marked nouns in Russian or Latin are rare, both in the lexicon and in number of occurrences. It can happen but it need not.

7. semantically unmarked interpretations are more frequent than semantically marked interpretations

This may well be true. Old material is often reduced and with the same number of references to different kind of things and parity of references to old and new the reduced expressions are more frequent. Also, Jakobson's unmarked meanings are simpler and it may well be that we use simple concepts more frequently than complex ones. If it is so, we can explain why phonological markedness seems in (weak) harmony with Jakobson semantic markedness: the connection is Zipf's law that associates frequent words with short forms.

My intermediate conclusion is that markedness should be qualified and that harmony between those qualified markednesses is not automatically given. I have no counterexamples against (10) only.

4 History increases the Unmarked?

It might be however that it is always too early to expect full harmony and that language history tends to harmony and linguistic evolution eliminates disharmonic cases. Maybe so, but also the opposite happens.

The counterexamples to (3) *Formally unmarked forms express semantically unmarked meanings* are all weak: the unmarked form may be as semantically marked as its marked alternative. They are due to the phenomenon of a optional marking turning into obligatory marking, so that a full opposition arises between the formally marked form and its formally unmarked alternative. It seems that a good evolutionary reconstruction can be given for the historical process that leads to this kind of oppositions.

Marked forms arise for marked meanings that need to be marked. But when they enter into opposition with the unmarked form, the unmarked form loses its unmarked meaning and obtains a marked meaning which expresses that the meaning of the marked form does not apply.

The exceptions to (1) *typologically marked is formally marked* may also be explained by historical processes. Reduced pronouns seem to arise by a phonological divergence between the accented and unaccented occurrences and subsequent cliticisation and morphemisation. This process creates typologically marked forms that are formally unmarked.

Similarly, the exception to principle (3) that typologically marked forms have typologically marked interpretations, also can be understood from a historical process. Reduced pronouns have topical meanings, because they derive from the unaccented use of the pronoun and because they could not be used for non-topical antecedents because they cannot bear contrastive accent.

I have another speculative but rather interesting puzzle, deriving from a possibility considered in [9]. She derives an analytic perfect with a non-nominative dative (ergativoid) subject from an earlier passive form in which the demoted agent appears in the dative. One can think of the following schematic development. In the first phase the passive form is recruited for theta-marking because in the passive there is a case distinction between the agent and the theme. This destroys the connection between the passive and the marking of the theme as the topic and leads to a reanalysis of the passive form as an active form with a perfective meaning. The new form may completely replace the old active form. Now look at two points in time: one (t_1) where the old active still occurs but has become rare and a second moment (t_2) -not much later- where it has completely disappeared. It seems we should be able to think of the two moments as parts of a continuous process. But at t_1 , the old active is the formally unmarked and the typologically unmarked form while it is probabilistically the marked form. The emerging new active that is really a passive is morphologically and typologically marked and probabilistically unmarked. At t_2 , it becomes the new active, is suddenly typologically unmarked and the morphologically least marked form. Also its meaning has changed from marked (theme is topic) to unmarked (agent is topic). The probabilistic point of view seems to capture the continuity, while the markedness terminology does not seem to deal with the continuity in the situation at all.

I conclude that history can break harmony as well as establish it. If it can establish it, Zipf's frequency law can be explained by phonological decay under frequency and would associate frequent words with shorter forms. If frequent concepts are the simpler ones, this would explain why phonological unmarkedness goes with semantic unmarkedness. It would put formal markedness in harmony with semantic markedness in the Jakobson sense. But it is not easy to make these explanations very precise and they cannot be given for all the harmony principles.

For me, the conclusion cannot be the one that a distinguished colleague at the workshop expressed: that we should eliminate the concept of markedness from linguistic theory as being hopelessly vague. I would just plead for adding a lot more adjectives to it when one uses it, perhaps even more than I have wanted to do in this note. Markedness refers to the exception and the rule and thereby to what distinguishes natural from artificial languages. The trouble is that there is a lot of normalities in language, each with its own exceptions. And as another distinguished colleague remarked, it is the central concept that connects synchronic and diachronic linguistics and both of these with typology. So markedness should not be given up, but it should also not be accepted without qualifications and harmonies should not be assumed without foundational argument. It is important to understand why two markedness dimensions should be in harmony and why, if they should be, there are still exceptions.

References

- [1] Judith Aissen. Markedness and subject choice in Optimality Theory. *Natural Language and Linguistic Theory*, 17:673–711, 1999.
- [2] Reinhard Blutner. Some aspects of optimality in natural language interpretation. *Journal of Semantics*, 17(3):189–216, 2000.
- [3] Joan Bresnan. The emergence of the unmarked pronoun. In J. Grimshaw G. Legendre and S. Vikner, editors, *Optimality-Theoretic Syntax*, pages 113–142. MIT press, Cambridge (MA), 2001.
- [4] William Croft. *Typology and Universals*. Cambridge University Press, Cambridge, 2002.
- [5] J. Gundel, N. Hedberg, and R. Zacharski. Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):274–307, 1993.
- [6] Petra Hendriks and Helen de Hoop. Optimality theoretic semantics. *Linguistics and Philosophy*, 24:1–32, 2001.
- [7] R.O. Jakobson. Zero sign. In Linda R. Waugh and Morris Halle, editors, *Roman Jakobson. Russian and Slavic grammar: Studies 1931-1981.*, pages 151–160. Mouton de Gruyter, Berlin, 1939/1984.
- [8] R.O. Jakobson. Quest for the essence of language. In Linda R. Waugh and Monique Monville-Burston, editors, *On Language-Roman Jakobson*, pages 407–421. Harvard Press, Cambridge (MA), 1965/1995.
- [9] Maia Sackokia. New analytical perfects in Modern Georgian. In H. Zeevat D. de Jongh and M. Nilsenova, editors, *Proceedings of the 3rd and 4th International Tbilisi Symposium on Language, Logic and Computation*, pages 1–10. ILLC/ICLC, Amsterdam/Tbilisi, 2002.
- [10] Robert van Rooy. Signalling games select Horn strategies. *Linguistics and Philosophy*, t.a.
- [11] Henk Zeevat. The asymmetry of optimality theoretic syntax and semantics. *Journal of Semantics*, 17(3):243–262, 2001.