

**Symposium on linguistic categorization and the nature of linguistic categories
University of Tromsø, 23-24 April 2009**

Philipp Conzett
Department of Language and Linguistics
University of Tromsø
NO-9037 Tromsø
philipp.conzett@uit.no

The nature of nominal classification: the case of grammatical gender¹

1. Introduction

In this talk I will approach the nature of linguistic categorisation with regard to nominal classification, especially grammatical gender. The two main questions to be addressed are stated in the abstract and repeated in (1) on the hand-out:

- (1) (a) What kind of linguistic category is gender²?
- (b) What is the nature of gender assignment rules, and what are they meant to account for?

2. A scale of classificatory techniques

As you can see in (2) ...

- (2) “The common characteristic of nominal classification systems is that they classify the nouns of a given language.” (Serzisko 1982:95)

The definition in (3) gives the minimal set of criteria every noun classification system has to fulfil.

- (3) Noun classification is (op.cit.:96):
 - (a) a grouping of all nouns of a language into a delimited number of classes
 - (b) so that there is at least some overt indication of the class of a noun within any sentence in which it occurs within one of a certain set of syntactic constructions,
 - (c) and this indication is not entirely within the noun-word.

Based on this definition Serzisko and others distinguish between the three main types of noun classificatory techniques listed in (4): Numeral classification, noun class, and gender.

- (4) Three main types of noun classificatory techniques (op. cit.):
 - (a) Numeral classification
 - (b) Noun class
 - (c) Gender

The primary feature of **numeral classification** is that this technique occurs in quantifying constructions. The class marker occurs with numerals, the question word for “how many”,

1 I would like to thank Endre Mørck and Trond Trosterud for useful comments on an earlier draft of this paper.

2 Throughout this paper *gender* is used in the restricted sense of *lexical gender* (cf. Dahl 2000).

and quantifiers like ‘some’, ‘many’, etc. You can see an example in (5).

- (5) Numeral classification in Malay (op. cit.:98):
se-batang sungai ‘1-long object river’

The constituting features of **noun class** systems are (1) explicit marking, for instance the classification is morphologically expressed on the noun; and (2) the classification is bound to number expression. For every class there is - in an ideal system - a singular and a non-identical plural affix. An example from Swahili is given in (6). In (a) you see the singular noun class prefix *ki*, and in (b) the corresponding plural noun class prefix *vi*.

- (6) Noun class in Swahili (loc. cit.):³
- (a) *ki*-kapu *ki*-zuri ni *ki*-pi, hi-*ki* *ki*-dogo au *ki*-le *ki*-kubwa
cl-basket *cl*-nice be *cl*-which this-*cl* *cl*-small or *cl*-that *cl*-big
‘Which is the nicest basket, the small one or the big one?’
- (b) *vi*-kapu *vi*-zuri ni *vi*-pi, hi-*vi* *vi*-dogo au *vi*-le *vi*-kubwa
‘Which are the nicest baskets, the small ones or the big ones?’

The classification in **gender** systems is typically expressed implicitly, which means that gender is not indicated on the noun itself, but through agreement in associated words (cf. Corbett 1991:1; Hockett 1958:231). The allocation of nouns to classes is fixed, that is nouns cannot change their gender only by a change of agreement. A further characteristic of gender systems is that the number of classes is restricted to maximally three. The examples in (7) illustrate the three gender system in Norwegian. So, *ein bil* is masculine, *ei lampe* is feminine, and *eit bord* is neuter.

- (7) Gender in Norwegian:⁴
- (a) Dette er *ein* bil. Han er *min*
this is a-MASC car it-MASC is mine-MASC
- (b) Dette er *ei* lampe. Ho er *mi*
 a-FEM lamp it-FEM mine-FEM
- (c) Dette er *eit* bord. Det er *mitt*
 a-NEUT table it-NEUT mine-NEUT

According to Serzisko (1982) the three techniques can be understood as focal instances on a scale which is determined by the three parameters in (8).

- (8) Three parameters of classificatory techniques (op. cit.:99):
- (a) Grammaticality
(b) Semanticity
(c) Variability

3 *cl* = noun class marker

4 If not otherwise stated, Norwegian is represented by the written standard variety of Nynorsk throughout the paper.

The indicator for **grammaticality** on the morphological level is the degree of boundedness between class marker and classified noun or other nominal categories. The degree of boundedness is highest in those cases where the class markers are fused with the noun or those other nominal categories.

As you can see from the italicised gender markers in (7), gender systems are highly grammatical.

As stated in (9), the parameter of **semanticity** comprises according to Serzisko (op. cit.:103) two aspects: (a) The semantic complexity of the classificatory system, and (b) the criteria for allocating nouns to classes. **Semantic complexity** means here simply the number of different classes. As you can see from the examples in (9a), in numeral classification systems we usually find a rather large number of classes. In noun class systems the number is considerably lower, and in gender systems, the number is restricted to at most 3.

(9) Two aspects of semanticity (op. cit.:103f.):

(a) Semantic complexity of classificatory system:

- Numeral classification: Lao 95 classifiers, Chinese 24, Thai 246, Burmese 189, Vietnamese 140
- Noun class: between 2 and 40
- Gender⁵: at most 3

(b) Criteria for allocating nouns to classes:

- Semantic
- Mechanic (i.e. arbitrary, by rote learning)

Alternative second aspect of semanticity (cf. Fischer 2004:39, Ibrahim 1973:52):

(b₂) Global class meaning

By the aspect in (9b) Serzisko refers to the degree to which class allocation is based on semantic criteria. As you can see in (10), according to Serzisko ...

(10) “[P]ure semantic classification is very rare and surely not found throughout the entire classificatory system. In most cases it is only a ‘core group’ of nouns which possess the relevant semantic features, while the other nouns have to be considered as exceptions.” (Op. cit.:106)

As we shall see in a minute, for gender systems there have been proposed other allocation criteria than semantic ones. Therefore I would like to propose an alternative second aspect of semanticity, namely **global class meaning**, as stated in (9b₂). That is: what is the common semantic property of the nouns which fall into the class? Global class meaning is usually clearly indicated and transparent in numeral classification. As you can see in (5), the class marker *batang* in Malay is highly lexical and obviously denotes the global class meaning. On the other hand, gender markers do not “mean” anything longer. This is probably the reason why the nouns in the different gender classes of a language usually do not constitute globally coherent semantic patterns. So, on a *global* level Ibrahim’s conclusion cited in (11) seems to prove correct: “[...] semantically, gender is an empty category”.

(11) “[...] semantically, gender is an empty category”. (Ibrahim 1973:52)

5 Some scholars do not distinguish between noun class and gender systems (e.g. Corbett 1991).

This means that a global class meaning is usually highly concealed or non-existent in gender systems. Noun class systems are usually placed somewhere in the middle of the scale for global class meaning transparency.

The last parameter listed in (8), **variability**, is defined by Serzisko (1982:108) as the possibility of classifying nouns temporarily.

The degree of variability may be fairly high in noun class systems, but it is even higher in numeral classification. This is shown in the quotation in (12) which illustrates the situation in Burmese, where ...

- (12) “Each classifier can be applied to several objects of similar attributes, and some of these objects may have more than one attribute. A person may be classed as a human being or as an animal according to his behaviour [...]; and a sword as just a straight thing or a weapon. The choice of the classifier is prompted by the occasion.” (Hla Pe 1965:170f.; cited in Serzisko 1982:110f.)

In gender systems, on the other hand, temporary classification is usually not possible or at least very much restricted. A noun belongs to one and only one gender class.

In (13) I have summarised the ranking of the three noun classificatory techniques on the scales of grammaticality, semanticity, and variability.

- (13) A scale of nominal classificatory techniques (adopted from op. cit.):

| | NUM. CLF. | NOUN CLASS | GENDER |
|---|--------------------------------|------------|--------------|
| morphological boundedness of class marker | independent (lexical) elements | affixes | fusion |
| GRAMMATICALITY | | | |
| | low | -----> | high |
| semantic complexity | ≅ 40 | ≅ 5 | ≤ 3 |
| global class meaning | transparent | | non-existent |
| SEMANTICITY | | | |
| | high | -----> | low |
| temporary classification | high | restricted | zero |
| VARIABILITY | | | |
| | high | -----> | low |

3. What kind of linguistic category is gender?

Based on the diagram in (13) we can now return to the question in (1a): What kind of linguistic category is gender? A preliminary answer is given in (14): ...

- (14) Gender is a classificatory feature of nouns that
- (a) is morphologically expressed by fused class markers on associated words
 - (b) is inherently tied to the noun
 - (c) has no coherent global class meaning

The last point is probably the reason why traditionally many scholars have considered gender in Indo-European languages to be an *arbitrary* feature. But as some linguists argue, this holds mostly for modern languages. At earlier stages of Indo-European, they claim, gender had clearly distinguished global class meaning, which they usually refer to as the *function* of gender. As you can infer from (15), not all scholars agree on what this original function of Indo-European gender was. The proposals include biological sex, animacy, and definiteness.

- (15) The function of gender at earlier stages of Indo-European⁶:
- (a) Biological sex: male vs. female (cf. Grimm 1967 [1831])
 - (b) Animacy: animate vs. inanimate (cf. Aikhenvald 2003)
 - (c) Individuation: individual vs. continuous (cf. Leiss 2000, Vogel 2000)
 - (d) Definiteness: definite vs. indefinite (cf. Ostrowski 1985)

Regardless of the original function, the Indo-European gender system has dispersed in the course of language evolution. The mechanisms behind this process is usually described as extension based on semantic and/or formal similarity, as stated in (16).

- (16) Dispersion of gender class meaning
- (a) by extension based on semantic similarity (metaphor, metonymy, etc.) and/or
 - (b) by extension based on formal similarity (phonological, morphological)

The result of this dispersion is the modern situation where most Indo-European gender systems do not have any overt and consistent class meaning on a *global* level any more. However, as Fischer points out in (17), if one renounces the claim of global scope, the function of gender in modern Indo-European languages could be described as contributing in distinguishing meaningful expressions.

- (17) “Lässt man den Anspruch auf Globalität fallen, kann man mit Werner (1975, 39) vorsichtig formulieren: Das Genus «ist beteiligt, bedeutungstragende Ausdrücke zu unterscheiden».” (Fischer 2004:82)

In other words, gender is not seen any more as a fundamental perspectivisation technique tied to some few overall semantic distinctions. Research on gender has therefore not been focusing on the functional aspect of gender classification. Much of the recent literature on gender deals rather with the question of gender assignment, that is how nouns get their gender.

⁶ The relationship between the notion of *global class meaning* and *function* is not quite clear. However, this issue will not be discussed here.

4. What is the nature of gender assignment rules?

This leads us back to the first part of the second question posed in (1): “What is the nature of gender assignment rules?”. Most scholars agree on that in most modern Indo-European languages, biological sex has become the most clearly visible semantic content of gender classes⁷, and that the gender assignment of animate nouns to a large extent is based on this semantic feature. However, apart from biological sex, gender assignment has been treated rather differently in the literature. The claims range all the way from totally arbitrary to essentially rule-based, as you see in (18).

- (18) Arbitrary gender assignment (e.g. Bloomfield 1933)
 Rule-based gender assignment (e.g. Corbett 1991)

Let us first consider some rule-based approaches. As we saw, by the processes in (16) old class meanings have got distorted. But by the same processes also new semantic groupings and clusters have arisen. In addition to semantic features, also formal characteristics have been suggested to be at work in gender assignment. In general, three basic types of gender assignment rules are usually assumed in the literature: semantic, morphological, and phonological, as listed in (19).

- (19) Three types of gender assignment rules (cf. op. cit.:7f.):
- (a) semantic: based on meaning
 - (b) morphological: based on word-structure (derivation and inflection)
 - (c) phonological: based on sound structure

Regularities between gender on the one hand and semantic, morphological or phonological properties of the noun on the other hand can be more or less strict. Regularities showing a weak or considerable degree of consistency are often referred to as tendencies, whereas more highly consistent regularities are often given the status of rules. The overview in (20) gives some examples of regularities with varying degrees of consistency.

- (20) Consistency of gender assignment regularities:⁸

| Type | Consistency | Language | Regularity | Example | Exception | Reference |
|----------------------------------|-------------|--------------------|---|---|--------------------------------|------------------------|
| (a) Semantic | +++++++ | Tamil | Nouns denoting male rationals (= gods or male humans) are masculine | <i>aan</i> 'man', <i>civan</i> 'Shiva' | | Corbett (1991:8f.) |
| (b) Semantic | ++++ | German | Nouns denoting superordinates are neuter | <i>Obst</i> 'fruit', <i>Ding</i> 'thing' | <i>Pflanze</i> F 'plant' | Zubin & Köpcke (1986) |
| (c) Morphological (inflectional) | +++++++ | Arapesh | Nouns of declensional class 1 have gender <i>i</i> | <i>agaby</i> 'back' | | Aronoff (1994:107f.) |
| (d) Morphological (inflectional) | ++++ | Swedish | Nouns with plural ending -Ø (= declension class 6) are neuter | <i>hus</i> 'house', <i>barn</i> 'child' | <i>lärare</i> C 'teacher' | Källström (1996:157f.) |
| (e) Morphological (derivational) | +++++++ | Danish | Nouns derived in -dom are of common gender («fælleskøn») | <i>sygdom</i> 'illness' | | Allan et al. (1995:23) |
| (f) Morphological (derivational) | ++++ | Norwegian (bokmål) | Nouns derived in -skap are neuter | <i>ekteskap</i> 'marriage' | <i>latskap</i> M 'laziness' | NRG (106f.) |
| (g) Phonological | +++++++ | Qafar | Nouns whose citation form ends in an accented vowel are feminine | <i>catô</i> 'help' | | Corbett (1991:51f.) |
| (h) Phonological | ++++ | French | Nouns whose citation form ends in /-ø/ are feminine | <i>nation</i> 'nation' | <i>mouton</i> M 'sheep' | Müller (2000:354) |

In Tamil e.g., we find some highly consistent semantic gender assignment regularities. In this language, nouns denoting male rationals are assigned masculine gender, whereas nouns

7 In some languages or dialects, gender has been entirely organised around the notion of semantic *boundedness*: The Danish dialect of West-Jutlandic has developed a two-gender system of its own: *en*-gender for count nouns, and *et*-gender for mass nouns (Perridon 2003:241).

8 C = common gender, F = feminine, M = masculine

denoting female rationals are feminine, leaving the neuter gender to nouns denoting non-rationals. These assignment rules are based on the semantic notions of biological sex and rationality. However, in other languages, regularities have been observed that are based on other semantic criteria. In German, nouns referring to superordinates are claimed to be predominantly neuter, but there are a number of exceptions. The other regularities shown in (20) can be described in a similar way.

The fact that many of the proposed gender assignment rules are only partially consistent means of course that for many nouns, one assignment rule is not enough in order for them to be assigned their gender. In other words, many nouns fall under the scope of more than one assignment rule or regularity. Let me give you some examples from Norwegian.

The probably most consistent semantic gender regularity in modern Norwegian is based on biological sex. From the shaded fields in the overview in (21) you can see that of all animate root nouns denoting males, 99 % are masculine, 95 % of all root nouns denoting females are feminine, and of all animate root nouns that do not specify the gender of their referents, 81% are masculine.

(21) Quantitative gender distribution for BIOLOGICAL SEX in Norwegian⁹:

| | | # | | | | % | | | | |
|-----------|-------------|------|------|------|------|----|----|----|-----|-----|
| | | f | m | n | Σ | f | m | n | Σ | Σ |
| ANIMATE | MALE | 2 | 351 | 0 | 353 | 1 | 99 | 0 | 100 | 15 |
| | FEMALE | 226 | 4 | 9 | 239 | 95 | 2 | 4 | 100 | 10 |
| | UNSPECIFIED | 196 | 1407 | 143 | 1746 | 11 | 81 | 8 | 100 | 75 |
| | Σ | 424 | 1762 | 152 | 2338 | 18 | 75 | 7 | 100 | 100 |
| INANIMATE | | 1288 | 3466 | 1524 | 6278 | 21 | 55 | 24 | 100 | 73 |
| Σ | | 1712 | 5228 | 1676 | 8616 | 20 | 61 | 19 | 100 | 100 |

Apart from biological sex, also other semantic regularities can be found in the Norwegian gender system. The table in (22) summarises the gender distribution for root nouns denoting long or longish objects. 58% of them are masculine.

(22) Quantitative gender distribution for LONG(ISH) in Norwegian:

| | # | | | | % | | | |
|-----------|-----|-----|----|-----|----|----|----|-----|
| | f | m | n | Σ | f | m | n | Σ |
| LONG(ISH) | 102 | 198 | 39 | 339 | 30 | 58 | 12 | 100 |

In addition, there are certain phonological properties that are more or less closely tied to a specific gender. In (23) you have a summary of the quantitative relationship between gender and two coda sequences in monosyllabic root nouns in Norwegian. With 71% the coda /-kt/ is most closely associated with feminine gender, while 65% of the nouns ending in /-st/ are masculine.

(23) Quantitative gender distribution for monosyllabic coda sequences in Norwegian:

| | # | | | | % | | | |
|---------|----|----|----|-----|----|----|----|-----|
| | f | m | n | Σ | f | m | n | Σ |
| /-kt/ | 36 | 8 | 7 | 51 | 71 | 16 | 14 | 100 |
| /-st/ | 24 | 71 | 14 | 109 | 22 | 65 | 13 | 100 |
| (/-ast/ | 4 | 5 | 3 | 12 | 33 | 42 | 25 | 100 |
| (/-ist/ | 5 | 13 | 0 | 18 | 28 | 72 | 0 | 100 |

9 If not indicated otherwise, the numbers for Nynorsk in this paper are based on all root nouns from *Nynorskordboka* (NOB) (“Dictionary of Nynorsk”). I would like to thank the editors at *Seksjon for leksikografi og målføregranskning*, University of Oslo, for granting me access to the electronic dictionary files, and Christian-Emil Ore at *Eining for digital dokumentasjon*, University of Oslo, for adjusting the data for computational use.

And finally, the overview in (24) shows the gender distribution for bisyllabic root nouns ending in unstressed /-e/. Also in this group feminine gender dominates with 62%.

(24) Quantitative gender distribution for bisyllabic unstressed /-e/ in Norwegian:

| | # | | | | % | | | |
|------|------|-----|-----|------|----|----|---|-----|
| | f | m | n | Σ | f | m | n | Σ |
| /-e/ | 1687 | 833 | 212 | 2732 | 62 | 30 | 8 | 100 |

The semantic regularities in (21) and (22), and the phonological regularities in (23) and (24) form the empirical basis of the gender assignment rules in (25).

(25) Gender assignment rules in Norwegian (cf. e.g. Trosterud 2001):

- (a) Semantic rule: Nouns denoting males and females are masculine and feminine respectively.
- (b) Semantic rule: Nouns denoting animates but without specifying biological sex are masculine.
- (c) Semantic rule: Nouns denoting long(ish) objects are masculine.
- (d) Phonological rule: Nouns ending in /-kt/ are feminine.
- (e) Phonological rule: Bisyllabic nouns ending in unstressed /-e/ are feminine.

In addition Trosterud proposes the semantic assignment rule in (26), that is ...

(26) Semantic assignment rule (Trosterud 2001:41):

Nouns denoting flakes, sheets and flat surfaces are neuter.¹⁰

Now, some nouns fall under the scope of two or more of the assignment rules in (25) and (26). As you see in (27), the nouns *knekt*, *sjikt*, *vakt*, and *lekt* all end in the coda /-kt/. Therefore the phonological assignment rule in (25d) predicts feminine gender for all four nouns. The noun *knekt* denotes male, and the noun *vakt* does not specify the biological sex of the person. Therefore the semantic assignment rules in (25a) and (b) predict masculine gender for both nouns. The noun *sjikt* falls under the scope of the semantic “flake and surface” rule in (26), and the noun *lekt* denotes a long object.

The rule conflicts are solved in different ways, as illustrated in (27). In (a) we see that semantics wins over phonology in the case of *knekt* and *sjikt*, whereas in (b) phonology wins over semantics in the case of *vakt* and *lekt*. Furthermore, the examples in (27) also show that rule conflicts can be resolved in favour of any of the three genders, masculine, feminine, and neuter.

(27) Conflict between semantic and phonological assignment rules

- (a) Semantics “wins over” phonology:
 - knekt* MASC ‘fellow, young man’
 - Semantic assignment: MALE --> MASC [(25a)]
 - Phonological assignment: monosyllabic /-kt/ --> FEM [(25d)]
 - sjikt* NEUT ‘layer’
 - Semantic assignment: FLAKE/SURFACE --> NEUT [(26)]
 - Phonological assignment: monosyllabic /-kt/ --> FEM [(25d)]

¹⁰ The quantitative gender distribution for root nouns with this semantic property is: F: 66 (21%); M: 171 (55%); N: 74 (24%). The empirical facts would thus indicate that this semantic property should be tied to masculine rather than neuter gender. In this case, both semantic and phonological assignment would fail for the noun *sjikt* NEUT; cf. (28).

- (b) Phonology “wins over” semantics:
vakt FEM ‘guard’
 Semantic assignment: UNSPEC. SEX --> MASC [(25b)]
 Phonological assignment: monosyllabic /-kt/ --> FEM [(25d)]
lekt FEM ‘lath’
 Semantic assignment: LONG(ISH) OBJECT --> MASC [(25c)]
 Phonological assignment: monosyllabic /-kt/ --> FEM [(25d)]
kjempe FEM ‘giant’
 Semantic assignment: MALE --> MASC [(25a)]
 Phonological assignment: bisyllabic /-e/ --> FEM [(25e)]

As you can see in (28), in some cases both semantic and phonological assignment rules fail to account for the gender of a noun. The two nouns *mast* and *list* denote both long objects and they both are monosyllabic ending in /-st/. Both properties should trigger masculine gender according to the assignment rules in (25) and the gender distribution in (23). However, both nouns are feminine.

- (28) Both semantics and phonology fail:
mast FEM ‘mast’; *list* FEM ‘moulding’
 Semantic assignment: LONG(ISH) OBJECT --> MASC [(25c)]
 Phonological assignment: monosyllabic /-st/ --> MASC [(23)]¹¹

The recent literature on gender assignment is of course aware of such cases of rule conflict and rule failure, and there have been proposed different strategies. Three of them are summarised in (29).

In (a), Corbett and Fraser claim that in case of conflicts between semantic and formal assignment rules, semantic rules take precedence. As we saw in (27b), examples like *vakt* and *lekt* show that Corbett and Fraser’s claim does not hold for Norwegian. It does not apply to Italian either, as shown by Thornton (2007), and I suppose that the same is true for other languages that have been studied in depth.

In (b), Nessel constraints Corbett and Fraser’s universal to apply to biological sex only. However, examples like the Norwegian noun *kjempe* in (27b) indicate that also biological sex is in some cases overridden by formal assignment rules. In (30) you see several cases where the biological sex rule is overridden by another semantic rule based on affectiveness.

And finally, in (29c), Rice leaves assignment conflicts to constraint ranking supplied by language specific general gender markedness hierarchies. In Norwegian neuter is, according to Rice, the most marked gender, feminine is less marked, and masculine is the least marked gender. In the case of rule conflicts, or in OT-terms: constraint conflicts, the rule assigning the least marked gender wins. But, as is obvious from (27), according to this model, nouns like *sjikt* should be assigned feminine gender, and nouns like *lekt* should be assigned masculine gender in Norwegian. And once more, Thornton reports similar cases from Italian.

- (29) Strategies for gender assignment rule conflicts (cf. Thornton 2007):
 (a) Semantic >> formal:
 “As is universally the case, the formal gender assignment rules [...] are dominated

¹¹ There have not been proposed any phonological assignment rules for monosyllabic root nouns in /-st/, but based on the gender distribution in (23), the most obvious one is of course “monosyllabic nouns in /-st/ are masculine”.

- by the semantic gender assignment rules” (Corbett & Fraser 2000:321).
- (b) The Core Semantic Override Principle:
“Rules referring to biological sex take precedence in gender assignment” (Nesset 2006:1386).
- (c) Markedness hierarchy in Optimality Theory:
E.g. Norwegian: *NEUT >> *FEM >> *MASC (Rice 2006:1400)

(30) BIOLOGICAL SEX overridden by other semantic assignment rule:

fruentimmer ‘women’, *ludder* ‘whore, slut’, *viv* ‘wife’ (all NEUT), *vamp* MASC ‘vamp’

The discussion so far on the nature of gender assignment can be summarised as in (31). There are few or no strict gender assignment rules. There are few or no universally valid strategies for rule conflicts. In addition, nouns like *mast* and *list* are not accounted for by strictly rule-based models.

- (31) The nature of gender assignment so far:
- (a) Few or no strict gender assignment rules
 - (b) Few or no universally valid rule conflict solution strategies
 - (c) Nouns like *mast* and *list* (28) are not accounted for by strictly rule-based models

Another example where Core Semantic Override Principle fails:

bølle fm ‘rå og brutal (valds)mann; ramp’

Here, both semantic and phonological assignment rules fail. In other words, we could say that “conventionality wins”.

5. What are gender assignment rules meant to account for?

At this point I would like to go back to the second part of the second question I asked at the very beginning in (1b): “What are gender assignment rules meant to account for?” Or in other words, as formulated by Thornton in (32): “Who assigns gender to what?”. She distinguishes between four cases: (a) a child learning L1, (b) a speaker learning L2, (c) a linguist writing a computational grammar of Lx, and (d) an adult speaker of L1. She also points out that, “[u]nfortunately, the literature on gender assignment often fails to distinguish between the different cases in [(32)], and the rules that are proposed are meant to work for all these different purposes” (op. cit.:187”).

| | | | | |
|------|--|-----|---|---------------------|
| (32) | <i>Who assigns gender?</i> | | <i>To what?</i> | (Thornton 2007:186) |
| (a) | Child learning L1 | --- | To all nouns in L1 | |
| (b) | Speaker learning L2 | --- | To all nouns in L2 | |
| (c) | Linguist writing a computational grammar of Lx | --- | To all nouns in Lx | |
| (d) | Adult speaker of L1 | --- | To nouns in L1 that control gender agreement on targets but do not (yet) have a gender feature in their lexical entry | |

(33) “Unfortunately, the literature on gender assignment often fails to distinguish between the different cases in [(32)], and the rules that are proposed are meant to work for all these different purposes” (op. cit.:187”).

In the reference work on gender, Corbett, here cited in (34), assumes that one of the main reasons why gender assignment has to be rule-based is that “[n]ative speakers typically make few or no mistakes in the use of gender; if the gender of every noun were remembered individually, we would expect more errors” (Corbett 1991: 7).

(34) “[N]ative speakers typically make few or no mistakes in the use of gender; if the gender of every noun were remembered individually, we would expect more errors” (Corbett 1991: 7).

Such arguments are now seen as outdated, and as van Berkum argues in (35), “given the weak reliability of most assignment rules proposed in the literature, and the fact that several conflicting rules may apply to a single word, one would rather expect computation to yield the highest error rate.” (van Berkum 1996: 42)

(35) “One would not want to argue, for example, that the low incidence of word-form errors is evidence that speakers compute the form of tens of thousands of words from their meaning. If people can store the essentially arbitrary form of tens of thousands of words, why wouldn't they be able to store gender as well? Actually, given the weak reliability of most assignment rules proposed in the literature, and the fact that several conflicting rules may apply to a single word, one would rather expect computation to yield the highest error rate.” (van Berkum 1996: 42)

So, as indicated in (36), from a psycholinguistic point of view, as far as language production by adult speakers of L1 is concerned, gender *storage* is preferable to strictly rule-based gender *assignment*. Nouns that are not accounted for by assignment rules have to be listed in what is

- (a) *pipe* ‘pipe’ (ON¹² *pípa*) (c) *pinne* ‘stick’ (ON *pinni*)
pølse ‘sausage’ (ON **pylsa*) *stolpe* ‘pole’ (ON *stolpi*)
- (b) *kølle* ‘club’ (< Danish)
søyle ‘pillar’ (< German)

However, the nouns *pinne* and *stolpe* in (c) are not problematic in the network model, because just like the nouns in (a), they bear their gender as an inherent property, so they do not need to be assigned gender by any rule system any more. But as with any other lexical property, e.g. word meaning, the gender of a noun can change during language history.

As an example we can look at the development of some weak masculine nouns in Norwegian. Weak nouns in Old Norse ended either in unstressed /-a/ or /-i/. The unstressed /-i/ was mostly found in masculine nouns, and to some lesser extent in neuter nouns. The unstressed /-a/ was almost exclusively found in feminine nouns and was therefore a kind of noun internal feminine gender marker. In many modern varieties of Norwegian both /-a/ and /-i/ have been reduced to unstressed /-e/.

(39) Cases of gender development of bisyllabic nouns in unstressed /-e/ Norwegian:¹³

| | | | | |
|-----|--|---|--|---------------------------------------|
| (a) | Old Norse MASC <i>hani</i> <i>oxi</i> | > | modern Norwegian MASC <i>hane</i> ‘rooster’ <i>okse</i> ‘bull’ | motivation for MASC MALE <--> MASC |
| (b) | MASC <i>pinni</i> <i>stolpi</i> | > | MASC(/FEM) ¹⁴ <i>pinne</i> ‘stick’ <i>stolpe</i> ‘pole’ | LONG(ISH) <--> MASC |
| (c) | MASC <i>posi</i> <i>bolli</i> | > | MASC(/FEM) ¹⁵ <i>pose</i> ‘bag’ <i>bolle</i> ‘bowl’ | ? |
| (d) | MASC <i>lampi</i> <i>pakki</i> | > | FEM/MASC <i>lampe</i> ‘lamp’ <i>pakke</i> ‘packet’ | ? |
| (e) | MASC <i>api</i> <i>krabbi</i> | > | FEM/MASC <i>ape</i> ‘monkey’ <i>krabbe</i> ‘crab’ | UNSPEC. SEX <--> MASC |

The nouns you see in (39) are all bisyllabic ending in unstressed /-e/. As illustrated in (b) to (e), some of them have shifted from masculine to feminine in some varieties of Norwegian. This shift is most probably due to the strong correlation between unstressed /-e/ and feminine gender. In these varieties, it can be argued that unstressed /-e/ is becoming a new feminine gender marker.

As you also see in (39), in some varieties, masculine is retained in some or all of these nouns. In some cases one could argue that masculine gender has “survived” because the noun

12 ON = Old Norse

13 Similarly dispersed patterns and developments in varieties of Norwegian are also found in other areas of the gender system discussed in this paper. However, for the sake of simplicity, they have not been discussed here; cf. footnote 2 above.

14 In the written standard only masculine.

15 In the written standard only masculine.

still provides a semantic motivation for masculine. For the examples *ape* and *krabbe* in (e) the motivation could be that they both denote animates, and the masculine gender of *pinne* and *stolpe* in (b) could be motivated by the longish shape of their referents. However, as the nouns *lampe* and *pakke* in (d), and *pose* and *bolle* in (c) show, in many cases there is no apparent reason for masculine. So, maybe we should just regard masculine gender in (b) to (e) as conventional.

Finally, we see in (e) that the correlation between gender and biological sex is still more entrenched than the unstressed /-e/-correlation in virtually all varieties of Norwegian. So, *hane* and *okse* have still masculine gender.¹⁶

In general, I would say that a trivial but important feature of gender is that nouns keep their gender unless there are strong reasons for change. The examples in (39) suggest that the inclination to gender change is a gradient property which is probably based on differences in usage frequency or other factors that yet have to be investigated in detail.

A similar gradience phenomenon is also found in the initial phase of gender assignment to new nouns. Some new nouns may fit clearly into an existing pattern in the network, whereas other vacillate between different gender patterns. An example is given in (40). The noun *abstract* is neuter for some speakers of Norwegian, while other speakers prefer masculine. Still others are unsure about what gender to assign to this noun.

(40) Gender vacillation in initial phase of gender assignment in Norwegian:

e.g. *abstract* ‘abstract, summary’

(a) N: *sammandrag*, *utkast*; flakes and sheets-rule (26, cf. *ark* ‘(sheet of) paper’, *papir* ‘paper’)

(b) M: globally most entrenched gender

There are several possibilities of semantic motivation for neuter gender. We have the flakes and sheets-rule from (26), but also near-synonyms and related nouns like *sammandrag* and *utkast*, which both are neuter. The most probable motivation for masculine is its status as the globally most entrenched gender in modern Norwegian. The choice of gender for individual speakers depends therefore on how familiar they are with this word, and how entrenched the patterns in (40a) and (b) are in their mental lexicon or grammar.

To summarise, we can say that the exemplar-based gender model outlined in this talk accounts for both the development of existing nouns as well as the integration of new nouns. The gradient nature of gender connections ranging from motivated to conventionalised is neatly reflected in both gender continuity, gender change, and gender vacillation in the initial phase of network integration.

Further study of gender systems in different languages and different language varieties will contribute to our understanding of noun classification in particular, but also of the nature of linguistic categorisation at large.

16 There are though dialectal exceptions here. In some Norwegian varieties, we find feminine forms like *den viltre oksa* (‘the wild ox’; young female newsreader from Southern Norway on TV 2 sports news 20. April 2009).

References

- Aikhenvald, Alexandra Y. (2003): *Classifiers. A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.
- Allan, Robin; Holmes, Philip & Lundskaer-Nielsen, Tom (1995): *Danish: a comprehensive grammar*. London and New York: Routledge.
- Aronoff, Mark (1994): *Morphology by Itself*. Cambridge and London: The MIT Press.
- Berkum, Johannes Josephus Augustinus van (1996): *The psycholinguistics of grammatical gender: Studies in language comprehension and production*. Doctoral Dissertation, Max Planck Institute for Psycholinguistics. Nijmegen: Nijmegen University Press.
- Bloomfield, Leonard (1933): *Language*. New York: Holt, Rinehart and Winston.
- Conzett, Philipp (2006) Gender assignment and the structure of the lexicon. *STUF Sprachtypologie und Universalienforschung* 59 (3), 223-240.
- Corbett, Greville G. (1991): *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. & Norman M. Fraser (2000): Gender assignment: a typology and a model, in: Gunter Senft (ed.), *Systems of nominal classification*. Cambridge: Cambridge University Press, 293-325.
- Dahl, Östen (2000): Animacy and the notion of semantic gender, in: Unterbeck et al., 99-115.
- Fischer, Rudolf Josef (2004): *Genuszuordnung. Theorie und Praxis am Beispiel des Deutschen*. Frankfurt am Main: Peter Lang.
- Grimm, Jacob (1967): *Deutsche Grammatik III*. Herausgegeben von Gustav Roethe und Edward Schröder (Documenta linguistica Grammatiken des 19. Jahrhunderts Reihe 5), reprint of the edition from 1890, original from 1831. Hildesheim: Georg Olms Verlagsbuchhandlung.
- Hla Pe (1965): A re-examination of Burmese 'Classifiers', *Lingua* 15, 163-185.
- Hockett, Charles F. (1958): *A course in modern linguistics*. New York: MacMillan.
- Ibrahim, Muhammad Hassan (1973): *Grammatical gender. Its origin and development*. The Hague: Mouton.
- Källström, Roger (1996): On gender assignment in Swedish, in: Kjartan G. Ottósson, Ruth V. Fjeld & Arne Torp (eds.), *The Nordic Languages and Modern Linguistics* 9, 151–167.
- Langacker, Ronald W. (1987): *Foundations of Cognitive Grammar, vol. 1*. Stanford, California: Stanford University Press.
- Leiss, Elisabeth (2000): Gender in Old High German, in: Unterbeck et al., 237-258.
- Müller, Natascha (2000): Gender and number in acquisition, in: Unterbeck et al., 351–399.
- Nesset, Tore (2006): Gender meets the Usage-Based Model: Four Principles of Rule Interaction in Gender Assignment. *Lingua* 116, 1369-1393.
- NRG = Faarlund, Jan Terje; Lie, Svein & Vannebo, Kjell Ivar (1997): *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Ostrowski, Manfred (1985): Zur Entstehung und Entwicklung des indogermanischen Neutrums, in: Bernfried Schlerath (ed.), *Grammatische Kategorien: Funktion und Geschichte* (Akten der VII. Fachtagung der Indogermanischen Gesellschaft Berlin, 20.-25. Februar 1983). Wiesbaden: Reichert, 313-323.
- Perridon, Harry (2003): Language contact and grammatical change: the case of Bergen;

Tijdschrift voor Skandinavistiek vol. 24 (2). URL
<http://dpc.uba.uva.nl/tvs/vol24/nr02/art05>

Rice, Curt (2006): Optimizing gender. *Lingua* 116, 1394-1417.

Salmons, Joe (1993): The structure of the lexicon: evidence from German gender assignment. *Studies in Language* 17-2, 411–435.

Serzisko, Fritz (1982): GENDER, NOUN CLASS, and NUMERAL CLASSIFICATION: a scale of classificatory techniques, in: René Dirven & Günter Radden (eds.), *Issues in the Theory of Universal Grammar* (Tübinger Beiträge zur Linguistik 196). Tübingen: Gunter Narr, 95-123.

Thornton, Anna (2007): Constraining Gender Assignment Rules, in: Geert Booij Geert, Luca Ducceschi, Bernard Fradin, Emiliano Guevara, Angela Ralli, & Sergio Scalise (eds.), *Online Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)* Fréjus 15-18 September 2005, University of Bologna, 185-208. URL <http://mmm.lingue.unibo.it/>

Trosterud, Trond (2001): Genustilordning i norsk er regelstyrt, *Norsk Lingvistisk Tidsskrift* 19, 29–58.

Unterbeck, Barbara, Matti Rissanen, Terttu Nevalainen & Mirja Saari (eds.) (2000): *Gender in Grammar and Cognition*. Berlin/New York: Mouton de Gruyter.

Vogel, Petra Maria (2000): Nominal abstracts and gender in Modern German: A “quantitative” approach towards the function of gender, in: Unterbeck et al., 461-493.

Werner, Otmar (1975): Zum Genus im Deutschen. *Deutsche Sprache* 1/1975, 35-58.

Zubin, David & Köpcke, Klaus Michael (1986): Gender and Folk Taxonomy: The Indexical Relation Between Grammatical and Lexical Categorization, in: Craig, Colette (ed.), *Noun Classes and Categorization*. Amsterdam and Philadelphia: John Benjamins, 139–180.