

PHILIPP CONZETT (Tromsø)

## Gender assignment and the structure of the lexicon\*

### Abstract

This article provides a general survey of a number of (mostly Indo-European) gender systems. In recent years, regularities found in those systems have been accounted for by rule-based approaches where gender is assigned on-line by symbolic rules. A critical investigation of those accounts suggests that gender of existing nouns is stored individually, but that there must also be a mechanism that assigns gender to new nouns. It is argued that this behaviour is more adequately accounted for in a network model and it is shown how gender assignment can shed light on lexical structure at large.

### 1. Introduction

In recent years, theoretical interest in the issue of gender assignment has increased. For instance, CORBETT (1991: 8) emphasizes that “[f]rom a theoretical point of view, assignment systems have important implications for attempts to determine the structure of the lexicon”. In the same work, he argues that gender assignment is not arbitrary but rule-based (CORBETT 1991, e.g. chapter 4), sketches a number of gender assignment rules for many languages based on regularities found in their systems (CORBETT 1991, chapters 2 and 3), and addresses the theoretical implications of the assumption that gender assignment is rule-based (CORBETT 1991).

In this paper, a closer look is taken at how different theoretical approaches to gender assignment embedded in the wider field of lexical representation and processing can shed light on the structure of the lexicon and its status in a theory of language. Section 2 surveys different kinds of gender assignment regularities in the languages of the world, especially in Indo-European, looks at different models of rule-based gender assignment which make use of these regularities, and at their implications for our understanding of the lexicon. Section 3 provides a critical discussion of the psycholinguistic evidence on gender assignment and what it can tell us about the validity of the models previously discussed. The aim of the final section 4 is to see what the results from sections 2 and 3 imply for the structure of the lexicon in general and its status in linguistic theory.

\* For discussion of various issues in this paper, I am grateful to GIANNOULA GIANNOULOPOULOU, MARIT JULIEN, TORSTEN LEUSCHNER, MICHELE MIOZZO, TORE NESSET, TROND TROSTERUD and MARIT R. WESTERGAARD. Responsibility for errors of all kinds remains with the author.

## 2. Gender assignment regularities and rule-based theories

### 2.1. Types and consistency of assignment regularities

It used to be assumed that gender assignment is largely arbitrary, i.e. that gender is stored individually and independently for each noun in the lexicon. One of the main reasons why this view has been largely abandoned is the fact that closer examination reveals a great deal of regularity in gender assignment systems, in the sense of structured patterns of correlation between the gender of nouns and other inherent properties of the same nouns. Depending on the kind of property in question, we can distinguish two main types of gender assignment regularities, viz. semantic and formal, as shown in (1). Semantic regularities state consistent correlations between gender and meaning, formal regularities state consistent correlations between gender and formal properties. Information about the form

	Type	Consistency	Language	Regularity	Examples	Exceptions	Reference
a.	semantic	+++++++	Tamil	nouns denoting male rationals (=gods or male humans) are masculine	<i>aan</i> 'man', <i>civanN</i> 'Shiva'		CORBETT (1991: 8f.)
b.	semantic	++++	German	nouns denoting superordinates are neuter	<i>Obst</i> 'fruit', <i>Ding</i> 'thing'	<i>Pflanze</i> 'plant'	ZUBIN & KÖPCKE (1986)
c.	morpho-logical (inflectional)	+++++++	Arapesh	nouns of declensional class 1 have gender I	<i>agaby</i> 'back'		ARONOFF (1994: 107f.)
d.	morpho-logical (inflectional)	+++++	Swedish	nouns with plural ending -Ø (=declensional class 6) are neuter	<i>hus</i> 'house', <i>barn</i> 'child'	<i>lärare</i> 'teacher'	KÄLLSTRÖM (1996: 157f.)
e.	morpho-logical (derivational)	+++++++	Danish	nouns derived in <i>-dom</i> are of common gender («fælleskøn»)	<i>sygdom</i> 'illness'		ALLAN et al. (1995: 23)
f.	morpho-logical (derivational)	++++	Norwegian (Bokmål)	nouns derived in <i>-skap</i> are neuter	<i>ekteskap</i> 'marriage'	<i>latskap</i> 'laziness'	NRG (106f.)
g.	phono-logical	+++++++	Qafar	nouns whose citation form ends in an accented vowel are feminine	<i>catò</i> 'help'		CORBETT (1991: 51f.)
h.	phono-logical	+++++	French	nouns whose citation form ends in /-ø/ are feminine	<i>nation</i> 'nation'	<i>mouton</i> 'sheep'	MÜLLER (2000: 354)

Table 1: Consistency of gender assignment regularities

can in turn be of two types, indicating either word-structure, comprising derivation and inflection (i.e. morphology), or sound-structure (i.e. phonology).

- (1) Types of gender assignment regularities [cf. CORBETT 1991: 7f.]  
 semantic: based on meaning  
 formal: morphological: based on word-structure (derivation, inflection)<sup>1</sup>  
 phonological: based on sound structure

Furthermore, the regularities can be more or less strict. Regularities showing a more or less strong degree of consistency are often referred to as tendencies, whereas more highly consistent regularities are often given the status of rules. The overview in Table 1 on the previous page gives some examples of regularities with varying degrees of consistency. In Tamil e.g., we find some highly consistent semantic gender assignment regularities: nouns

	Gender system type							
	cn		mf		mfn		Σ all types	
gender assignment regularities	#	%	#	%	#	%	#	%
S	2	40	4	15	1	3	7	10
Mi	0	0	2	8	8	22	10	15
Md	0	0	0	0	1	3	1	1
P	0	0	4	15	0	0	4	6
Σ «strict» types	2	40	10	38	10	28	22	33
S+M+P	2	40	3	12	3	8	8	12
S+M	1	20	9	35	21	58	31	46
S+P	0	0	3	12	2	6	5	7
M+P	0	0	1	4	0	0	1	1
Σ «mixed» types	3	60	16	62	26	72	45	67
Σ all types	5	100	26	100	36	100	67	100

M morphological      Md derivational      S semantic      f feminine  
 Mi inflectional      P phonological      c common gender      m masculine

Table 2: Assignment regularities in Indo-European gender systems<sup>2</sup>

<sup>1</sup> Morphological regularities can also be seen as forming types of their own, given that they are based on both semantic and formal information. Thanks to TORE NESSET for reminding me of this.

<sup>2</sup> The study was carried out by the author as a part of the Gender Project at the Center for Advanced Study in Theoretical Linguistics (CASTL) of the University of Tromsø, <<http://uit.no/castl/2447/4>>. The results were obtained primarily by browsing reference works. The overview is based on “non-anaphoric” gender, i.e. gender distinctions revealed in attributive and/or predicative agreement. The terms “common”, “feminine”, “masculine” and “neuter” are chosen for historical reasons; it would also have been possible to operate with numbers only.

denoting male rationals are assigned masculine gender, whereas nouns denoting female rationals are feminine, leaving the neuter gender for nouns denoting non-rationals. These assignment rules are based on the semantic notions of sex and rationality. In other languages, regularities may be based on other semantic criteria. In German, nouns referring to superordinates are predominantly neuter, but there are a number of exceptions. The other regularities shown in Table 1 can be described similarly.

The gender assignment systems of some languages seem to allow for semantic regularities only. As an example, CORBETT (1991: 8f.) cites Tamil where the whole system can be described in terms of rules like the one in (a) in Table 1. In most other languages, however, we can observe both semantic and formal regularities. A preliminary, but comprehensive, study of gender assignment regularities found in Indo-European languages can give us an impression of how prevalent the different kinds of regularities are. From the overview in Table 2, you can see that out of 67 languages with gender assignment regularities, 45 languages, i.e. about 67 per cent, allow for more than one type of assignment.

## 2.2. Rule-based accounts

Central to most recent literature on gender assignment is the claim that gender assignment systems not only show regularities, but that the assignment is essentially rule-based. The assignment rules are often stated in an SPE-like format (cf. e.g. CORBETT 1991, TROSTERUD 2001, ZUBIN & KÖPCKE 1986), but there are also attempts to implement the regular behaviour of gender assignment within the constraint-based framework of Optimality Theory (OT) (cf. RICE 2006). The main challenge for both approaches is how to handle the problem of rule (or constraint) interaction, i.e. how to account for what happens when a noun falls into the domain of two or more conflicting assignment rules (or constraints). The main idea behind these models is that rules interact in a structured way and are thus able to produce the gender for basically all nouns of a given language.

The following examples may give some idea of how rule interaction is handled in a rule-based theory. To illustrate how rule interaction is handled in OT, let us consider the German noun *Gemüse* 'vegetables'. According to RICE (2006), three specific assignment constraints are relevant to the gender assignment of this noun. The specific constraints are shown in (2). They are based on observed gender regularities in German.

- (2) Gender assignment constraints in German (adapted from RICE 2006: 1406):
- (2a) Nouns ending in *-e* must not be masculine or neuter  
(*<* phonological rule: nouns ending in *-e* are feminine)
  - (2b) Nouns derived in *Ge-* must not be masculine or feminine  
(*<* derivational rule: nouns derived in *Ge-* are neuter)
  - (2c) Nouns denoting superordinates must not be masculine or feminine  
(*<* semantic rule: nouns denoting superordinates are neuter)

In addition, RICE also postulates the general gender assignment hierarchy (3) which holds for the whole gender system of German. The hierarchy states that masculine is the default gender, i.e. the least marked gender, whereas neuter is the most marked gender in German.

- (3) General gender assignment hierarchy in German (RICE 2006: 1405):  
 \*neuter >> \*feminine >> \*masculine

Table 3 demonstrates how (2) combined with (3) can account for the correct gender assignment of the German noun *Gemüse*. Note that the specific constraints (2a) to (2c) are unranked, which means that their violations are evaluated as a whole. In the case of German *Gemüse*, the general gender hierarchy doesn't contribute to the gender assignment at all. However, in the case of the German noun *Pflanze* 'plant', the constraint conflict is not settled until we move on down to the general gender hierarchy. The optimal gender assignment of the noun is given in Table 4.

<i>Gemüse</i>	GENDER FEATURES			GENERAL GENDER HIERARCHY		
	*-e ⇒ m, n (2a)	*Ge- ⇒ m, f (2b)	*SUP ⇒ m, f (2c)	*NEUTER (3)	*FEMININE (3)	*MASCULINE (3)
a. <i>der Gemüse</i>	*	*!	*			*
b. <i>die Gemüse</i>		*	*!		*	
☞ c. <i>das Gemüse</i>	*			*		

Table 3: Optimal gender assignment of German *Gemüse* (RICE 2006: 1408)

<i>Pflanze</i>	GENDER FEATURES			GENERAL GENDER HIERARCHY		
	*-e ⇒ m, n (2a)	*Ge- ⇒ m, f (2b)	*SUP ⇒ m, f (2c)	*NEUTER (3)	*FEMININE (3)	*MASCULINE (3)
a. <i>der Pflanze</i>	*		*!			*
b. <i>die Pflanze</i>			*		*	
☞ c. <i>das Pflanze</i>	*			*!		

Table 4: Optimal gender assignment of German *Pflanze* (RICE 2006: 1408)

CORBETT's (1991: 40f.) model for Russian gender assignment is based on the ordered rules in (4):

- (4) Gender assignment rules for Russian [CORBETT 1991: 40f.]<sup>3</sup>  
 Semantic assignment
- (4a) nouns denoting males are masculine
- (4b) nouns denoting females are feminine  
 Morphological assignment  
 For declinable nouns:
- (4c<sub>d</sub>) nouns of declensional type I are masculine
- (4d<sub>d</sub>) nouns of declensional types II and III are feminine
- (4e<sub>d</sub>) nouns of declensional type IV are neuter

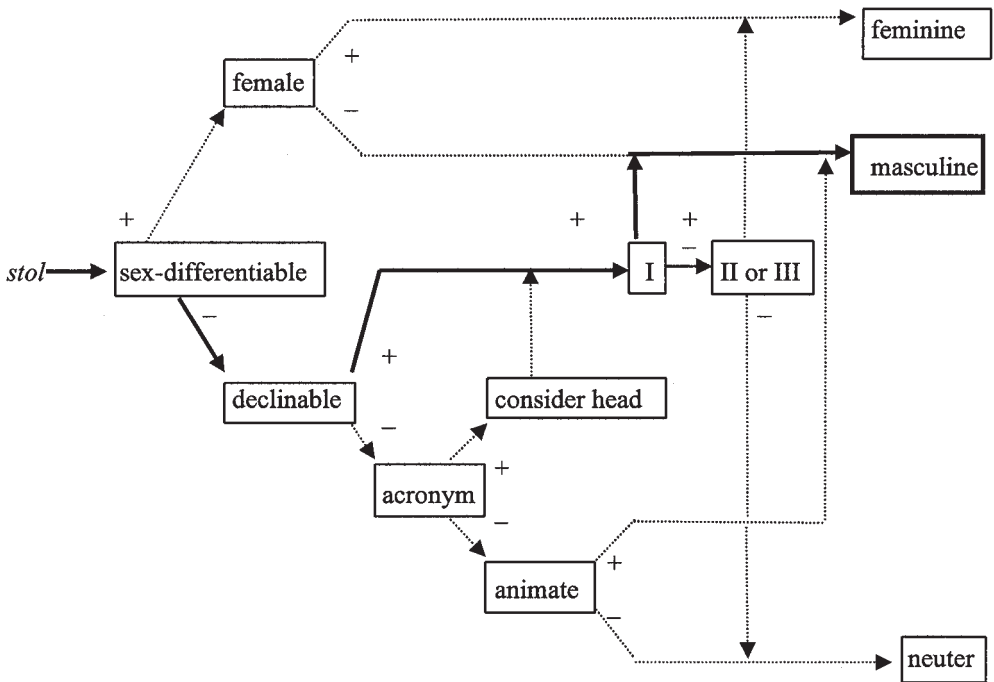
<sup>3</sup> CORBETT's rules are specified in a more complicated manner than the rules underlying constraints in an OT-approach. Especially, some of them (4c<sub>i</sub>–4e<sub>i</sub>) only apply within the domain of other rules (those referring to indeclinable nouns). It would have been more elegant to state these kinds of rules in their fully-fledged form, leaving the rule ordering to the Elsewhere Condition.

For indeclinable nouns:

- (4c<sub>i</sub>) for acronyms, take the head noun; the gender is then determined according to the morphological rules just given in (4c<sub>d</sub>) to (4e<sub>d</sub>)
- (4d<sub>i</sub>) nouns denoting animates are masculine
- (4e<sub>i</sub>) others are neuter

According to CORBETT, a Russian noun like *stol* 'chair' receives its correct gender by following the rules in (4) in the way illustrated by the flow chart in (5).

- (5) Gender assignment procedure for Russian *stol* (adapted from CORBETT 1991: 41)



### 2.3. Lexical structure in a rule-based model

For the current discussion it is not necessary to explore the mechanisms of the described models in more detail: by now, the main implications of a rule-based view on gender assignment, as held by CORBETT, should have become clear. In nearly all theories concerning rule-based gender assignment, the assignment of the correct gender to a noun of a given language depends crucially on other information about the noun. Furthermore, these theories claim that gender assignment is in principle rule-based for all nouns. CORBETT is aware that the observed assignment rules are not necessarily equally straightforward. Nevertheless, he points out that “in those languages which have been studied in depth, the gender of at least 85 per cent of the nouns can be predicted from information required independently in the lexicon” (CORBETT 1991: 68). Furthermore, CORBETT claims that gender assignment takes place in the lexicon (CORBETT 1991: 14).

To summarise, I have listed in (6) the main implications of a rule-based view of gender assignment as proposed by CORBETT and others for the structure of the lexicon in a theory of language:

- (6) Gender assignment in a rule-based model
- (6a) The lexical items we usually refer to as nouns are specified for:
- i. syntactic category: N
  - ii. inherent semantic characteristics of the referent
  - iii. phonological form (of the stem)
  - iv. morphological structure (derivational information)
  - v. inflection class membership
- (6b) The lexicon contains a system/module with gender assignment rules.
- (6c) The assignment rules are organised in such a way that each noun is assigned its gender.
- (6d) The assignment rules require the information stated in (a).

Most researchers on gender assignment seem to be conscious of the fact that their linguistic models of rule-based gender assignment are in the first instance no more than “black boxes” that assign the correct gender to any unassigned noun that happens to be inserted into the model (cf. CORBETT 1991: 68, KÖPCKE 1982: 133f.). Ultimately, the most interesting question to me seems to be whether the observed rules and/or regularities are part of native speakers’ mental assignment systems. As CORBETT (1991: 68f.) notes, answering this question will require careful research by linguists and psycholinguists, and he is convinced that this research will deepen our understanding of the structure of the lexicon, with important consequences for linguistic theory in turn. This means that before we accept the conclusions in (6), it is appropriate to carry out a critical investigation into the theoretical status of gender assignment. That is the topic of section 3.

### 3. Psycholinguistic evidence

#### 3.1. *Storing vs. computing*

CORBETT (1991) lists both linguistic and psycholinguistic evidence for the validity of the assignment models he sketches for several languages. The most self-evident argument for his rule-based account is the high amount of regularity found in the gender assignment systems discussed. According to CORBETT (1991: 70), people assign gender to nouns with ease, and they most plausibly do so by quite simply taking advantage of these regularities. Native speakers hardly ever make mistakes in gender assignment. From this CORBETT (1991: 7) concludes that “if the gender of every noun were remembered individually, we would expect more errors”. CORBETT thus argues that rule-based gender assignment is not only psychologically real but actually used in on-line language production. His main evidence is based on the assumption that stored information is much less reliable than information that is processed by a rule system based on other kinds of stored information. A similar view is expressed even more bluntly by ZUBIN & KÖPCKE (1981: 447): “[t]o have completely arbitrary gender assignment for the tens of thousands of nouns in

the average educated speaker's lexicon would present an insurmountable task to the language learner".

The argument put forward by CORBETT and ZUBIN & KÖPCKE seems to be underpinned by the widespread belief that the optimal linguistic model is the one with the least redundancy (as depicted by JACKENDOFF 2002: 30). But they do not give us any reason why storage would be less reliable and why it could not be handled by the language user. General reasoning leads VAN BERKUM to conclude that the opposite view is at least as plausible. As he points out, the low incidence of word-form errors is not usually taken as evidence that speakers compute the form of tens of thousands of vocabulary items from their meaning. He goes on to ask why people would not be able to store gender together with the "essentially arbitrary" form of the words. Furthermore, as we saw above in Table 1, many assignment rules proposed in the literature are only weakly reliable, and often several conflicting rules may apply to a single noun (see e.g. Tables 3 and 4). From these facts he concludes that "one would rather expect *computation* to yield the highest error rate" (VAN BERKUM 1996: 42). Furthermore, although we saw in section 2 that the proposed models are able to resolve conflicting rules, it does not seem very likely that this kind of rule-based processing would be more reliable in on-line production than storing. Therefore, VAN BERKUM's argument does indeed seem to be the more plausible one.

### 3.2. Psycholinguistic evidence from Italian

For our discussion, the interesting question now is, of course, whether there is any experimental evidence in favour of the view that gender is stored individually per noun. A number of psycholinguistic experiments, amongst others with native speakers of Italian, do indeed provide such evidence. Gender in Italian is highly interrelated to inflection class. The relation between these two categories is summarised in Table 5.

<b>Gender</b>	<b>Class 1</b> sg. -o, pl. -i	<b>Class 2</b> sg. -a, pl. -e	<b>Class 3</b> sg. -e, pl. -i	<b>Class 4</b> sg. -a, pl. -i	<b>Class 5</b> sg. -o, pl. -a	<b>Class 6</b> invariable
masculine	<i>libro</i> 'book'	–	<i>padre</i> 'father'	<i>papa</i> 'pope'	<i>uovo</i> 'egg'	<i>bar</i> 'coffee shop'
feminine	<i>mano</i> 'hand'	<i>casa</i> 'house'	<i>madre</i> 'mother'	<i>ala</i> 'wing'	<i>uova</i> 'eggs'	<i>star</i> 'famous person'

Table 5: Correlations between gender and inflection class in Italian (THORNTON 2001: 485)

We notice that it is not possible to predict either gender from declension class or vice versa.<sup>4</sup> As THORNTON (2001: 479) points out, "[t]he relationship between gender and inflection class, and the question of which of these two features determines the other on in a given language, has been the topic of much debate" and can only be answered language-specifically. For Italian, she demonstrates convincingly that information about gender is needed earlier by native speakers and is thus prior to inflection class. The main evidence is that for Italian nouns, "occurrence in the plural is much rarer than occurrence in a

<sup>4</sup> The only exception is class 2. If one knows that a noun belongs to this class, one also knows that it has feminine gender.



syntactic context which requires gender specification” (THORNTON 2001: 481). This again means that gender on its part has to be assigned based on other lexical information, e.g. semantic and/or phonological properties of the noun.<sup>5</sup>

Some semantic gender assignment rules for Italian nouns denoting animates, mostly based on sex distinctions, have been put forward. However, most rule-based accounts of Italian gender assignment emphasize the high significance of the singular vowel suffix for the prediction of gender. There is considerable agreement that there are phonological<sup>6</sup> assignment rules stating that nouns ending in the suffix *-o* and *-a* are masculine and feminine, respectively (cf. FERRARI-BRIDGERS 2002, MAIDEN & ROBUSTELLI 2000: 34f., MARCANTONIO & PRETTO 1991: 316f.). As for nouns ending in *-e*, on the other hand, correct gender assignment is not possible on the basis of formal information only.

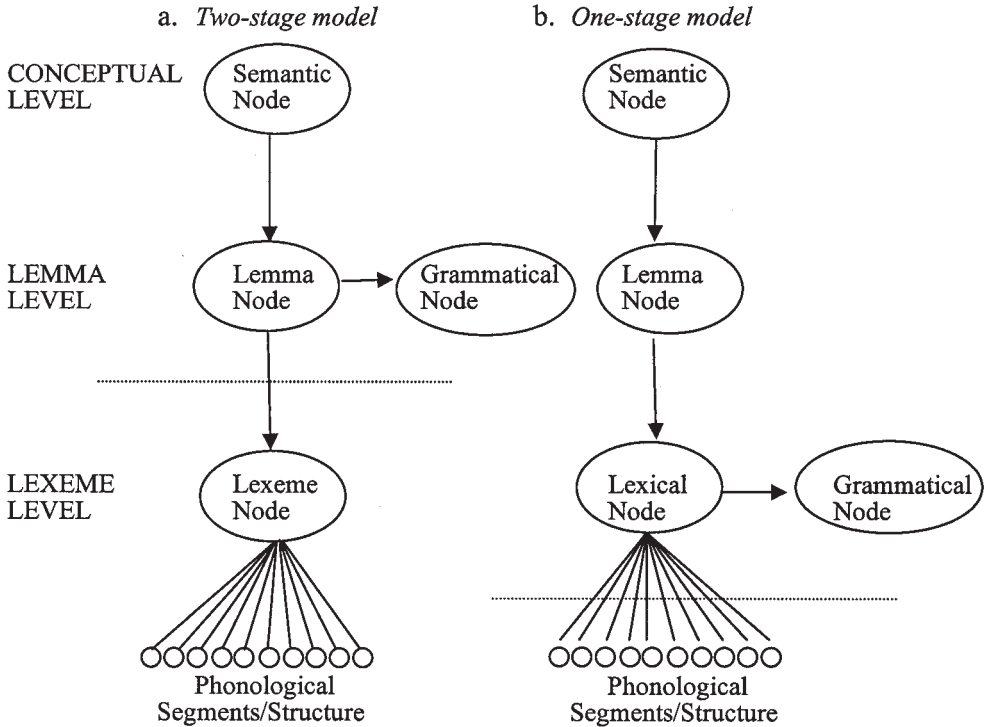
Let us now return to the psycholinguistic experiments. Can they provide evidence that gender is stored individually per noun? The experiments I will discuss here (BADECKER et al. 1995, MIOZZO & CARAMAZZA 1997, CARAMAZZA & MIOZZO 1997) were all designed to find out more about how native speakers of Italian retrieve information from the lexicon in on-line language production. The starting point for all experiments is the two-stage theory of lexical retrieval. The main claim of this theory is that the retrieval of lexical semantics, i.e. lemma retrieval on this account, is dissociated from the retrieval of lexical form in production: the first stage, lexical selection, maps non-verbal conceptual representations into modality neutral lemmas; lexical forms are only retrieved on the basis of the lemmas at a separate second stage. The evidence marshalled by BADECKER et al. (1995) comes from experiments with anomic patients. The fact that these patients are able to provide “definitional phrases” and pantomimic gestures even though they cannot provide any information as to the phonology or orthography of the target word is taken as an indication of the described dissociation in lexical retrieval. Schematically, the process can be illustrated as in (7a) below.

The problem with such experiments is that the subjects only provide semantic information about the target words (BADECKER et al. 1995: 194). The two-stage account assumes that the semantic information that the patients use when describing the target is accessed at a word-specific level, i.e. at the level of the lemma and not at the conceptual level. However, this is not the only way to interpret anomic behaviour. In an alternative model, which is illustrated in (7b), there is only one lexical level. On this account, in order to produce a particular word, only a single and unique record needs to be accessed, providing all the lexical information (semantic, grammatical, phonological and orthographic).

<sup>5</sup> Derived nouns are not included in the following discussion. As will be seen, for the point to be made, derivational information can be subsumed under semantic and phonological specification.

<sup>6</sup> If we accept THORNTON’s claim that gender is prior to inflection class, the basic form endings, viz. *-o*, *-a* and *-e* as discussed below, cannot be regarded as carrying morphological, i.e. inflectional, information at this stage, hence the term “phonological” assignment rules. In section 4 it will be shown that THORNTON’s analysis supports the conclusions in this paper.

- (7) Two models of lexical retrieval (adapted from CARAMAZZA & MIOZZO 1997: 311–313):



In order to confirm the two-stage model, it would have to be shown that an amonic patient, in addition to semantic information, also has access to word-specific information that is not related to conceptual information (BADECKER et al. 1995: 193f.). According to BADECKER et al. (1995: 194), “the strongest candidate would be information concerning [...] grammatical features, and grammatical gender is a case in point”. They claim that morphosyntactic lexical features such as grammatical gender are not encoded in or otherwise immediately derivable from the conceptual representation of the entity referred to by the noun used in a utterance. While there are Italian nouns (e. g. *padre* ‘father’ and *madre* ‘mother’) whose gender is determined straightforwardly by their semantic properties, the assignment of gender to nouns like *tavolo* (‘table’, masc.) and *sedia* (‘chair’, fem.) is not predictable from their semantics (1995: 194f.). To demonstrate this, BADECKER et al. (1995: 195) refer to the case of synonyms. If one grants that *astro* (‘star’, masc.) and *stella* (‘star’, fem.) are equally coherent lexical choices for a native speaker of Italian when talking about those celestial bodies, then one must also grant that gender can only be assigned on the basis of the specific lexical item that he or she chooses to employ (BADECKER et al. 1995: 195).

Two comments are in order here. First, BADECKER et al.’s argument is based on the assumption that the only semantic property that gender can be predicted from is the distinction between male and female. But as we saw in section 2, the gender of some nouns

denoting inanimates can be accounted for by semantic assignment rules that are not based on sex in a number of languages. The reason why Italian linguists have rarely focused on these kinds of regularities is no doubt the strong relationship between a given noun's gender and its formal properties. Second, the argument also presupposes that words like *astro* and *stella* are exact synonyms. However, the existence of real synonymy in natural language is highly controversial. As a main principle in language acquisition, CLARK (e. g. 1987) proposes the Principle of Contrast: "Every two forms contrast in meaning" (1987: 2). From this principle it would follow that exact synonyms should be rare or even non-existent. According to CARSTAIRS-McCARTHY (1994: 737), "[a]mong psycholinguists this implication is generally accepted, even though there is disagreement about the Principle of Contrast as an explanation for it". These observations present a serious challenge to BADECKER et al.'s reasoning, hence we have to bear these objections in mind when interpreting the psycholinguistic results in the next subsection (cf. also VAN BERKUM 1996: 186, fn. 43).

### 3.3. Assignment of known nouns

BADECKER et al. (1995) tested their hypothesis about the two-stage model in experiments with an anomic native speaker of Italian. In some pilot testing, it appeared that the subject did not have access to any phonological information about a word when he was unable to name it. Amazingly, he was nevertheless able to distinguish the grammatical gender of the target noun. In four subsequent experiments, this ability was further examined. The first experiment consisted of a picture-naming task. The subject was presented with 344 pictures of objects. Nouns whose grammatical gender is consistently predictable from semantic information were excluded, e.g. *ragazzo* 'boy' and *ragazza* 'girl', as were nouns whose gender can invariably be derived from phonological form or morphological composition (e.g. nouns ending in *-one*). If there was a general tendency for a particular ending to trigger one of the two genders, equal numbers of masculine and feminine nouns with this ending were chosen. It was also made sure that nouns with different frequency were included. The experiment was conducted as follows. Each time the subject appeared to be in an anomic state, i.e. when he wasn't able to name the object, he was asked to provide information about 1) the semantics, 2) the gender, and 3) the phonology and orthography of the target noun. Probing for information on form, i. e. 3), was discontinued after 52 omission trials because it was obvious that the subject could not provide any information at all about the phonology or orthography of the target noun. However, he could give a comprehensive description of the semantic properties of the object, and he correctly answered virtually all questions regarding semantics. Table 6 summarises the subject's performance in gender identification for the naming omissions accompanied by correct semantic information.

The subject was able to provide correct gender information in 95 % of all naming omission trials. Further analysis of the results showed that there was no difference in the level of performance on more regular noun targets (i.e. masculine singular nouns ending in /-o/ and feminine singular nouns ending in /-a/) and less regular noun targets. According to the authors, the results of this experiment indicate that the subject is generally incapable of accessing any phonological or orthographic information while in an anomic state. Nonetheless, he was able to provide the gender of the "unavailable" target nouns. Further-

		#	%
<b>a. Naming responses:</b>	Correct	218	63
	Naming omission	111	32
	Error	15	5
	Total	344	100
<b>b. Gender identification:</b>			
	Correct	106	95
	Error	3	3
	Don't know	2	2
	Total (= #naming omissions)	111	100

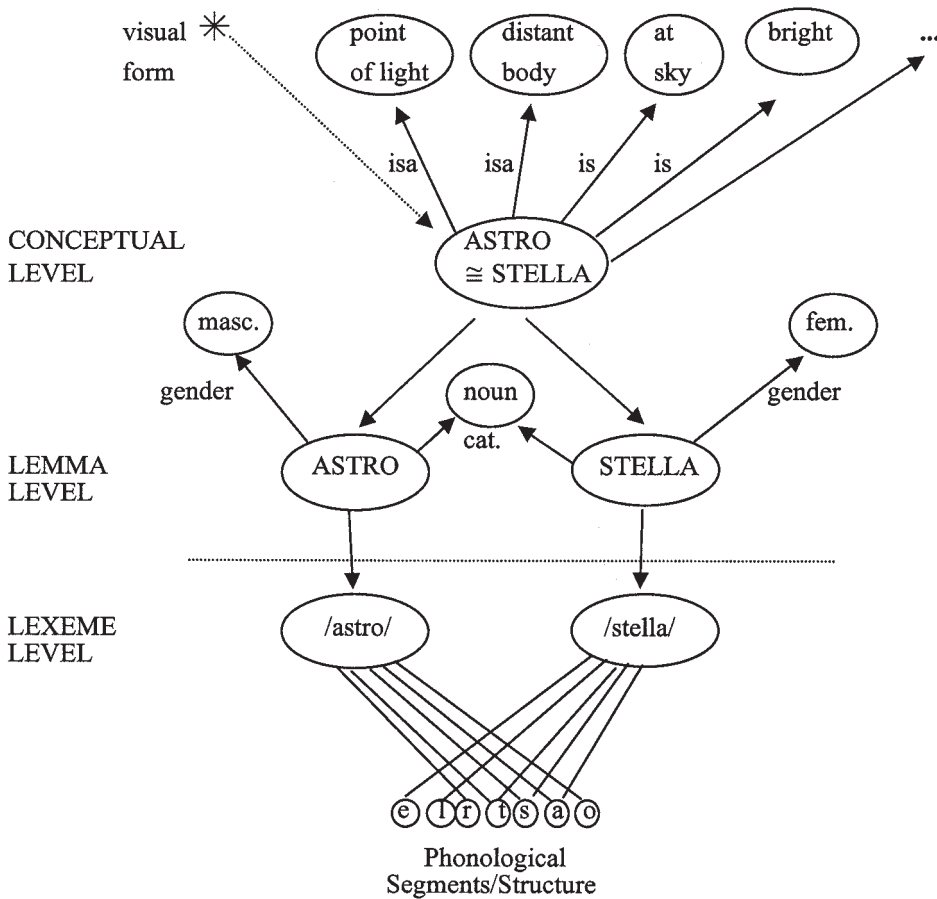
Table 6: Performance in gender identification in anomic state (BADECKER et al. 1995: 200)

more, they see no evidence that the subject's ability to specify the correct gender of a noun he could not name is effected by the regularity of the target nouns. Hence they interpret the results as strong support for the two-stage model of lexical production, which predicts "that the grammatical specification of a word is accessed independently of its phonological or orthographic form" (BADECKER et al. 1995: 201).<sup>7</sup>

BADECKER et al. (1995) conducted four more experiments in order to rule out any errors based on possibly questionable techniques used in the first experiment. The results of all four experiments reveal a gender identification rate of at least 95 % and thus seem to corroborate the hypothesis put forward in the two-stage model. Additional evidence comes from experiments on subjects in the tip-of-the-tongue state (e.g. CARAMAZZA & MIOZZO 1997, MIOZZO & CARAMAZZA 1997), which reveal basically the same pattern. CARAMAZZA & MIOZZO (1997) therefore propose the two-stage model of gender retrieval illustrated in (8).

<sup>7</sup> BADECKER et al. (1995) do not mention how other lexical features like morphological structure (derivational information) and information about inflection class could determine the gender of a noun, but apparently assume that this kind of information is also dependent on access to phonology. No matter whether the assignment rules in Italian mentioned above (i.e. nouns ending in the suffix *-o* and *-a* are masculine and feminine, respectively) are regarded as **morphological** or **phonological** (as on THORNTON's account, cf. fn. 6), they are in any case crucially dependent on the **phonological** information about the ending.

- (8) Two-stage model of gender retrieval in the “tip-of-the-tongue” state (adapted from CARAMAZZA & MIOZZO 1997):



The psycholinguistic evidence thus seems to confirm that native speakers of Italian are able in on-line production to produce the gender of nouns already known to them without having access to the phonology of the noun (stem+suffix). As indicated in the two objections to BADECKER et al. at the end of the preceding subsection one could also interpret the psycholinguistic results referred to above in favour of a purely semantic account of gender assignment. However, I am not aware of any rule-based account that manages to assign correct gender to at least 95 % of nouns like the ones used in the experiments described above without referring to formal properties. This strongly suggests that on-line gender production cannot solely be dependent on rule-based assignment, as claimed by CORBETT and others (cf. 3.1. above).

### 3.4. *Assignment of new nouns*

Let us now return to the evidence discussed by CORBETT (1991) and focus briefly on two of his other arguments which, as far as I can see, deal ultimately with the same phenomenon. The issue at stake here is how nouns are assigned gender that did not previously exist in the language of a community and, ultimately, of the individual. First, consider borrowings. It is obvious that new nouns borrowed into a language with gender have to be assigned a gender. It has been claimed that this assignment works according to the principle of semantic analogy with existing nouns, but CORBETT (1991: 70–82) argues convincingly that loan words get assigned their gender in accordance with the regularities described in section 2. However, a huge influx of loan words can have considerable consequences for the gender system of a language. For instance, if many loan words have the same formal or semantic properties, and these properties are associated with only one gender in the receiving language, then these regularities can gain in strength. More evidence comes from newly created or nonsense nouns. As CORBETT (1991: 7) notes, “when presented with invented words, speakers give them a gender and they do so with a high degree of consistency”. The psycholinguistic experiments conducted on this matter confirm that this kind of gender assignment is in accordance with the regularities found in the existing gender system of the language (1991: 89–92).

Limitations of space do not allow me to go into further detail, but the conclusion from these two pieces of evidence is that gender languages must allow not only for the storage of the gender of known nouns, they must also have a mechanism for the assignment of gender to new nouns. Furthermore, this mechanism seems to be based on the same regularities as the ones described in section 2. Despite this insight, we are still left with a couple of open questions. Some of them are summarised in (9):

- (9) Some open questions about gender assignment
  - (9a) How do children acquire gender systems? What can the study of child acquisition of gender tell us about gender systems in general?
  - (9b) How is gender stored/represented?
  - (9c) Is gender of existing nouns in normal language production (i.e. no language impairment or temporary naming omission and the like being involved) computed based on rules or regularities or just retrieved from storage or both retrieved and computed?

For reasons of space, I will not dwell on these questions here. Rather, it may be sufficient to note for the time being that the literature on these questions seems to suggest a distinction between the acquisition of gender (both in children and for new words) and the on-line production of gender (for existing words).

## 4. **Gender assignment and the structure of the lexicon**

What do the observations described in section 2 and the critical investigation into the psycholinguistic status of gender assignment in section 3 reveal about the status and the structure of the lexicon of adult speakers? It seems that there are in principle two basic approaches to satisfy the requirements that gender imposes on a theory of language. In

more traditional theory, gender assignment is based on symbolic rules, whereas in cognitive theory, gender assignment is accounted for by analogy with connections between different linguistic items. Both theories will need to provide a “place” or manner for storing lexical items and for a way to store and/or organise those items which we normally refer to as nouns such that the information in (10) is apparent in one way or another:

- (10) Lexical representation of nouns
- (10a) syntactic category: N
- (10b) gender
- (10c) inherent semantic characteristics of the referent
- (10d) phonological form
- (10e) inflection class membership

In a rule-based account, the information about a given noun’s morphological structure (i. e. derivational information) is not apparent from the stored item itself, but is provided by a set of symbolic derivational rules (perhaps in a separate morphology component of the grammar). In a network model as proposed by BYBEE (1985, 1988, 1995 a, 1995 b, 1998 and elsewhere), the morphological structure of lexical items is apparent from the way they are associated with each other. On this account, a new word is assigned its gender based on how it fits into the network of connections between existing words. It may well be the case that only the strongest associations are at work in the beginning and that associations are established after a while that were not apparent in the initial phase. Uncertainty and instability as to the gender of new nouns are richly attested in the literature (cf. CORBETT 1991: 75).

The network model appeals to me more than a model relying on symbolic rules. It seems unlikely that symbolic rules can be established without some stored information being present to serve as the basis for regularisation (LANGACKER 1987: 29, 46 and elsewhere). Since the psycholinguistic evidence discussed in section 3 strongly suggests that gender is stored individually per noun, and since the regular patterns in a structured network allow the speaker to assign gender to new words, why should we want to postulate separate symbolic rules? It seems implausible for gender to be assigned on-line based on rules that are in turn based on other rules, e.g. rules of word formation. The conclusions from section 3 are also supported by THORNTON’s analysis of the Italian noun system. THORNTON (2001: 484f.) stresses the importance of distinguishing between generalizations across the entire Italian lexicon on the one hand and productive gender assignment rules on the other hand. Productive gender assignment rules thus need to account only for nouns that are newly integrated into the lexicon. This is exactly how we would expect the network model to assign gender. Another, more general advantage of a network model is that it has no need for the arbitrary distinction between regular versus irregular behaviour. In a network, connections may differ in strength. The structure and strength of these relations is influenced by frequency and cognitive salience. The more frequently an item is realised, the more strengthened or entrenched the involved relations will be. The same holds for cognitive salience: items with high cognitive salience have a deeper impact on the network and thus also on other items. In short, a highly regular relationship between two patterns in a network will be due to a strong network connection (or strong network connections), whereas looser or less entrenched connections result in less regular patterns. I am not

aware of any gender systems in the literature whose behaviour could not be explained in a principled way by a network model as well as by a model based on symbolic rules (as proposed by PINKER 2000); in fact, suitable network models have already been put forward (e. g. by NESSET 2006). Further support for a network account is provided by computational models for language learning. Based on the rule set proposed by TROSTERUD (2001), HALSE (2004) uses different systems for rule induction in order to simulate how gender is assigned in Norwegian, a language in which all three types of gender assignment regularities (semantic, morphological, phonological) are represented. Moreover, the semantic regularities proposed in the literature include several tens of criteria,<sup>8</sup> and many of the suggested rules are only moderately consistent. Throughout HALSE's experiments (2004), it is the memory-based learner who achieves the lowest error rates of all three computational learning systems in assigning gender to new nouns.

If the network view is right, it implies a much more salient status in linguistic theory for the component usually called the lexicon. The main characteristics of a gender assignment system embedded in this kind of model of lexical representation and processing are summarised in (11).

- (11) Gender assignment in a network model of lexical representation and processing:<sup>9</sup>
- (11a) The lexical items we usually refer to as nouns are stored in such a way that the following information is apparent:
- i. syntactic category: N
  - ii. gender
  - iii. inherent semantic characteristics of the referent
  - iv. phonological form
  - v. morphological structure (derivational information)
  - vi. inflection class membership
- (11b) The lexical items are connected based on the properties described in (a).
- (11c) Gender is assigned to new nouns based on how they are integrated into the regularities emerging from the connections described in (b).

The aim of this paper has been to show how research on gender assignment – a topic that used to be considered a matter of arbitrary lexical listing – can shed new light on how language at large is represented and processed. More thorough research into the details of noun systems in the languages of the world will hopefully reveal more about gender assignment and its implementation in and interaction with the lexicon.

## References

- ALLAN, ROBIN; HOLMES, PHILIP & LUNDSKÆR-NIELSEN, TOM (1995): *Danish: a comprehensive grammar*. London & New York: Routledge.
- ARONOFF, MARK (1994): *Morphology by itself*. Cambridge and London: The MIT Press.
- BADECKER, WILLIAM; MIOZZO, MICHELE & ZANUTTINI, RAFFAELLA (1995): The two-stage model of lexical

<sup>8</sup> The number is based on browsing reference works and other literature on gender assignment in Norwegian.

<sup>9</sup> Cf. also SALMONS (1993).



- retrieval: evidence from a case of anomia with selective preservation of grammatical gender, in: *Cognition* 57, 193–216.
- BYBEE, JOAN L. (1985): *Morphology*. Amsterdam & Philadelphia: Benjamins.
- BYBEE, JOAN L. (1988): Morphology as lexical organization, in: HAMMOND, MICHAEL & NOONAN, MICHAEL (eds.), *Theoretical morphology*. New York: Academic Press, 119–141.
- BYBEE, JOAN L. (1995a): Diachronic and typological properties of morphology and their implications for representation, in: FELDMAN, LAURIE BETH (ed.), *Morphological aspects of language processing*. Hillsdale (New Jersey): Erlbaum, 225–246.
- BYBEE, JOAN L. (1995b): Regular morphology and the lexicon, in: *Language and cognitive processes* 10, 425–455.
- BYBEE, JOAN L. (1998): The emergent lexicon, in: *Chicago Linguistic Society* 34, 421–435.
- CARAMAZZA, ALFONSO & MIOZZO, MICHELE (1997): The relation between syntactic and phonological knowledge in lexical access: evidence from the ‘tip-of-the-tongue’ phenomenon, in: *Cognition* 64, 309–343.
- CARSTAIRS-MCCARTHY, ANDREW (1994): Inflection classes, gender, and the Principle of Contrast, in: *Language* 70.4, 737–788.
- CLARK, EVE V. (1987): The Principle of Contrast: a constraint on language acquisition, in: MAC WHINNEY, BRIAN (ed.), *Mechanisms of language acquisition*. Hillsdale, New Jersey: Erlbaum, 1–33.
- CORBETT, GREVILLE G. (1991): *Gender*. Cambridge: Cambridge University Press.
- FERRARI-BRIDGERS, FRANCA (2002): Gender arbitrariness in Italian? Handout distributed at the conference *The grammar of gender*, University of Oslo, November 28–29, 2002.
- HALSE, GRO (2004): *Genustilordning i nynorsk: Ei datamaskinell etterprøving*, [unpublished master thesis, Section for Computational Linguistics, University of Bergen].
- JACKENDOFF, RAY (2002): What’s in the lexicon?, in: NOOTEBOOM, STEB et al. (eds.), *Storage and computation in the language faculty*. Dordrecht, Boston & London: Kluwer, 23–58.
- KÄLLSTRÖM, ROGER (1996): On gender assignment in Swedish, in: *The Nordic Languages and Modern Linguistics* 9, 151–167.
- KÖPCKE, KLAUS-MICHAEL (1982): *Untersuchungen zum Genussystem der deutschen Gegenwartssprache*. Tübingen: Max Niemeyer.
- LANGACKER, RONALD W. (1987): *Foundations of Cognitive Grammar, vol. 1: theoretical prerequisites*. Stanford, California: Stanford University Press.
- MAIDEN, MARTIN & ROBUSTELLI, CECILIA (2000): *A reference grammar of Modern Italian*. Lincolnwood, Illinois: NTC Publishing Group.
- MARCANTONIO, ANGELA & PRETTO, ANNA MARIA (1991): Il nome, in: *Grande grammatica italiana di consultazione*. Bologna: Il Mulino, 315–332.
- MIOZZO, MICHELE & CARAMAZZA, ALFONSO (1997): Retrieval of lexical-syntactic features in tip-of-the-tongue states, in: *Journal of experimental psychology* 23.6, 1410–1423.
- MÜLLER, NATASCHA (2000): Gender and number in acquisition, in: UNTERBECK, BARBARA & RISSANEN, MATTI (eds.), *Gender in grammar and cognition*. Berlin: Mouton de Gruyter, 351–399.
- NESSET, TORE (2006): Gender meets the usage-based model: four principles of rule interaction in gender assignment, in: *Lingua* 116, 1369–1393.
- NRG = FAARLUND, JAN TERJE; LIE, SVEIN & VANNEBO, KJELL IVAR (1997): *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- PINKER, STEVEN (2000): *Words and rules*. New York: Perennial.
- RICE, CURT (2006): Optimizing gender, in: *Lingua* 116, 1394–1417.
- SALMONS, JOE (1993): The structure of the lexicon: evidence from German gender assignment, in: *Studies in Language* 17.2, 411–435.
- THORNTON, ANNA M. (2001): Some reflections on gender and inflectional class assignment in Italian, in: SCHANER-WOLLES, CHRIS et al. (eds.), *Naturally! Linguistic studies in honour of Wolfgang Ulrich Dressler presented on the occasion of his 60<sup>th</sup> birthday*. Torino: Rosenberg & Sellier, 479–487.
- TROSTERUD, TROND (2001): Genustilordning i norsk er regelstyrt, in: *Norsk Lingvistisk Tidsskrift* 19, 29–58.
- VAN BERKUM, JOHANNES JOSEPHUS AUGUSTINUS (1996): *The psycholinguistics of grammatical gender: studies in language comprehension and production*. Doctoral Dissertation, Max Planck Institute for Psycholinguistics. Nijmegen: Nijmegen University Press.
- ZUBIN, DAVID & KÖPCKE, KLAUS MICHAEL (1981): Gender: a less than arbitrary grammatical category, in: *Chicago Linguistic Society* 7, 439–449.

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 & Philadelphia: Benjamins, 139–180.

Bibliography of source grammars available on request.

PHILIPP CONZETT  
Universitetet i Tromsø  
Institutt for språk og språkvitskap  
NO - 9037 Tromsø  
NORWAY  
philipp.conzett@hum.uit.no