Language Sciences 60 (2017) 53–68

Contents lists available at ScienceDirect

Language Sciences

journal homepage: www.elsevier.com/locate/langsci

Calibrating complexity: How complex is a gender system?

Jenny Audring

Leiden University, LUCL, P.N. van Eyckhof 2, 2311 BV Leiden, The Netherlands

ARTICLE INFO

Article history:
Available online 12 November 2016

Keywords:
Complexity
Gender
Canonical typology
Difficulty

ABSTRACT

Grammatical gender is a many-sided phenomenon, involving complex relations between semantics, morphology, phonology, and syntax. Yet, not all gender systems across the world are equally complex. This paper presents a way to assess the complexity of gender systems in natural languages, building on the typological data collected in Corbett (1991, 2013) and applying the insights from Canonical Typology (chiefly Corbett, 2006, 2012; Corbett and Fedden, 2015). The result is a typologically responsible evaluation of the ways in which grammatical gender can be more or less complex. The analysis provides a descriptive basis for the assessment of difficulty in acquisition and processing, avoiding the a priori assumption that "complex" should always equal "difficult". Moreover, the article is intended as a methodological contribution by demonstrating the set-up and use of a calibration tool for the complexity of a grammatical subsystem.

1. Introduction

Speakers have intuitions about language complexity, and so do linguists. Asked about languages with complex grammatical gender, most linguists would mention French or Swahili rather than Italian or English. However, the gender systems of Swahili and French differ markedly in the way they are complex, and English gender might be complex in yet other ways.

One of the languages with a proverbially complex gender system is German. In a famous monograph, Köpcke (1982) proposed a set of rules that account for the gender of 90% of the language's monosyllabic nouns. This was – and still is – hailed as a remarkable feat, since German gender was thought to be complex to the point of arbitrariness (Bloomfield, 1933: 271, Maratsos, 1979). Yet, Köpcke's model in itself is considerably complex: it involves a total number of 44 different rules.

Köpcke's study, however, only investigates a single dimension of complexity: it focusses entirely on the rules of gender assignment. Other central aspects of gender, especially agreement, are not considered. This leaves the question of how simple or complex German gender might be in other dimensions, and indeed what these dimensions are. How can the complexity of a gender system be described and evaluated in a typologically responsible way?

E-mail address: j.audring@hum.leidenuniv.nl.

http://dx.doi.org/10.1016/j.langsci.2016.09.003
0388-0001/

Following up on a preliminary attempt in Audring (2014) and drawing on the wealth of typological and theoretical work available, this paper develops a methodology to answer this question. The focus is on *absolute* rather than *relative complexity* (see Kusters, 2003; Miestamo, 2008 for details on the difference). That is, the objective of this paper is a descriptive evaluation of the system rather than an assessment of how difficult it is to produce, perceive, or acquire.[1] The main reason for this choice is that difficulty cannot be studied directly – the only observable parameters are error patterns and processing time. If the acquisition, production, or processing of a structure proves to be slow or error-prone, an explanation for this fact will require an informed description of the complexity of this structure. Providing such a description, as well as suggesting a methodology for complexity calibration that can be applied to other areas, is the objective of this paper.

2. A two-step approach to complexity calibration

In the following we will assess the complexity of a grammatical subsystem in two steps. The first step is to identify the dimensions of variation for which it makes sense to evaluate a system or compare it to another. The second step establishes for each dimension what it means to be more or less complex. The point where all dimensions are at their lowest complexity is established as a theoretical zero, from which the degree of complexity can be calibrated outwards.

The first step builds on cross-linguistic knowledge of the grammatical subsystem in question. The second step requires awareness of the fact that whether or not a phenomenon is perceived as complex might depend on the perspective. For example, an inflectional paradigm showing syncretism may be considered less complex since it involves fewer than the expected forms. To illustrate, Table (1) gives the paradigm of the German singular definite article, which has 12 cells, but contains only 6 different forms.

(1) German singular definite article

Case / Gender	NOM	GEN	DAT	ACC
M	*der*	*des*	*dem*	*den*
F	*die*	*der*	*der*	*die*
N	*das*	*des*	*dem*	*das*

Yet, the simplification in morphology does not carry over to the syntax – German syntax requires the full range of case/gender pairings. This has the effect that a smaller form paradigm has to map onto a larger function paradigm. Hence, simplification in one area may bring complexification in another. Cases such as this require principled and explicit decisions on how complexity is evaluated.

A two-step approach in which the dimensions of complexity are identified and a hypothetical minimum is established yields a theoretical space in which languages can be situated.[2] This methodology is borrowed from Canonical Typology (Corbett, 2006, 2012; Brown et al., 2013). We will go further and argue that the parameters developed in Canonical Typology for agreement, morphosyntactic features and, indeed, gender itself can be directly applied to calibrate complexity. However, since canonicity and complexity are different issues, a brief theoretical discussion is in order.[3]

3. Canonicity vs. complexity

Canonical Typology has been developed with the aim of mapping out the theoretical and typological variety of individual linguistic phenomena. At the center lies the “canonical core” which is the ideal instance of the phenomenon in question, the point in which all definitions agree (2a). From this point outwards, a number of properties are identified in which real instances can diverge from the canonical, resulting in more debatable cases (2b).

[1] Note, though, that a description of a complex system already entails a basic assessment of its difficulty, if only in defining the amount of what needs to be learned in order to master the system. Thanks to Ray Jackendoff (p.c.) for pointing this out.

[2] It is worth pointing out that the two steps as outlined allow for calibration, but not for quantification. In order to actually measure complexity, the zero point and the dimensions of variation would need to be complemented by a scale that tells us how far a phenomenon diverges from zero. Devising such a scale is a third step, which goes beyond the limits of this paper. An exception is the dimension of gender values, which involves straightforward counting.

[3] This paper is not the first attempt to link canonicity and complexity. For a different approach, see Walther (2011) and Sagot and Walther (2011).