Modeling the complexity of grammatical gender systems
Three case studies in synchronic and diachronic typology

Francesca Di Garbo

Stockholm University

21.06.2017
Outline

My approach to grammatical complexity

1 – Exploring grammatical complexity crosslinguistically: the case of gender

2 – The evolutionary dynamics of gender systems’ complexity

3 – Correlates of restructuring in Bantu gender systems

References
My approach to grammatical complexity

How I define complexity: the number of parts in a system or the length of its description

**absolute complexity** (Dahl 2004, 2011; Miestamo 2008).

How I investigate complexity crosslinguistically: based on functional domains

**local complexity** (Miestamo 2006, 2008).

How I approach intra- and inter-language complexity variability: language variation and change can be influenced by social factors. **Languages as complex adaptive systems** (Beckner et al. 2009; Maitz & Németh 2014).
Dimensions of gender complexity
(Audring 2014)

Complexity of values  = number of genders.
Complexity of assignment  = number and scope of gender assignment rules.
Complexity and amount of formal marking  = number of gender agreement targets, frequency of gender marking in discourse.
Aim:

- To identify relevant dimensions of gender complexity and implement them into a complexity metric.

Method:

- One linguistic area: Africa
- Intra- and intergenealogical typology

Data:

- 84 languages from 17 different genealogical units from Africa.
- Descriptive resources and consultation of language experts.
The language sample

Legend
- Bantu
- Berber
- Chadic
- Cushitic
- Dizoid
- E. Nilotic
- Hadza
- Khoe−Kwadi
- Kwa
- Kxa
- Mel
- N.C. Atlantic
- Sandawe
- Semitic
- S. Omotic
- T.N. Omotic
- Tuu

Scale approx 1:78,000,000
Sex–Based vs. Non–Sex–Based systems

![Map of Africa showing sex-based and non-sex-based systems](image)

- Sex-based, 48
- Non-sex-based, 36

Scale approx 1:78,000,000

Legend:
- Blue circles: Sex-based, 48
- Red circles: Non-sex-based, 36
## Proposed complexity metric

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Feature</th>
<th>Value</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>Number of gender values ((GV))</td>
<td>Two genders</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Three</td>
<td>(1/3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Four</td>
<td>(2/3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Five or more</td>
<td>1</td>
</tr>
<tr>
<td>Assignment rules</td>
<td>Nature of assignment rules ((AR))</td>
<td>Semantic</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Semantic and formal</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Manipulation of gender assignment triggered by number/countability ((M1))</td>
<td>Absent</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Present</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Manipulation of gender assignment triggered by size ((M2))</td>
<td>Absent</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Present</td>
<td>1</td>
</tr>
<tr>
<td>Form. marking</td>
<td>Number of gender indexing domains ((IND))</td>
<td>One</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Two</td>
<td>(1/3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Three</td>
<td>(2/3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Four or more</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Cumulative exponence of gender and number ((CUM))</td>
<td>Noncumulative</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Partially cumulative</td>
<td>(1/2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cumulative</td>
<td>1</td>
</tr>
</tbody>
</table>
Results

The gender systems of the sampled languages are generally associated with high complexity scores.

Genealogical homogeneity.

Figure 1: Distribution of the GCSs
Genealogical homogeneity and outliers

The Bantu languages

Table 1: Bantu CGS

<table>
<thead>
<tr>
<th>ISO</th>
<th>Language</th>
<th>GCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>baz</td>
<td>Tunen</td>
<td>0.95</td>
</tr>
<tr>
<td>bem</td>
<td>Bemba</td>
<td>1</td>
</tr>
<tr>
<td>⇒bipBila</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>bvx</td>
<td>Dibole</td>
<td>0.78</td>
</tr>
<tr>
<td>cgg</td>
<td>Chiga</td>
<td>1</td>
</tr>
<tr>
<td>eto</td>
<td>Eton</td>
<td>0.83</td>
</tr>
<tr>
<td>kik</td>
<td>Gikuyu</td>
<td>1</td>
</tr>
<tr>
<td>kki</td>
<td>Kagulu</td>
<td>1</td>
</tr>
<tr>
<td>ksf</td>
<td>Bafia</td>
<td>0.78</td>
</tr>
<tr>
<td>lea</td>
<td>Lega</td>
<td>1</td>
</tr>
<tr>
<td>⇒linLingala (Kinshasa)</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>lol</td>
<td>Mongo-Nkundu</td>
<td>1</td>
</tr>
<tr>
<td>mcp</td>
<td>Makaa</td>
<td>1</td>
</tr>
<tr>
<td>ndg</td>
<td>Ndengereko</td>
<td>1</td>
</tr>
<tr>
<td>nso</td>
<td>Sotho,Northern</td>
<td>0.83</td>
</tr>
<tr>
<td>nya</td>
<td>Chichewa</td>
<td>1</td>
</tr>
<tr>
<td>sna</td>
<td>Shona</td>
<td>1</td>
</tr>
<tr>
<td>ssw</td>
<td>Swati</td>
<td>0.83</td>
</tr>
<tr>
<td>swa</td>
<td>Swahili</td>
<td>1</td>
</tr>
<tr>
<td>toi</td>
<td>Tonga</td>
<td>1</td>
</tr>
<tr>
<td>tsn</td>
<td>Tswana</td>
<td>0.83</td>
</tr>
<tr>
<td>ven</td>
<td>Venda</td>
<td>1</td>
</tr>
<tr>
<td>zul</td>
<td>Zulu</td>
<td>0.83</td>
</tr>
</tbody>
</table>

  - Two genders (Animate vs. Inanimate)
  - Semantic, non-manipulable assignment rules
  - NP-internal agreement only
  - Spoken in a linguistic area at the crossroads between different language families: Bantu, Ubangi, Nilotic.

  - Two genders (Animate vs. Inanimate) (fossilized noun class marking on nouns)
  - Semantic, non-manipulable assignment rules
  - Agreement only on third person pronouns and verbs
  - Kinshasa Lingala is the direct descendant of the Bangala pidgin.
Summary

- High gender complexity, constant across genealogically related languages.
- Gender complexity outliers have peculiar sociolinguistic profiles (multilingualism, contact).
Next steps

- Integrating the diachronic dimension to the crosslinguistic study of gender complexity.
- Digging into the sociohistorical factors that may contribute to complexification/simplification of gender systems.
The evolution of gender systems’ complexity
(project funded by the Wenner-Gren Foundations, Feb 2015 – Dec 2016)

Aims:
▶ To study the life-cycle of grammatical gender systems: emergence, expansion, reduction, loss.
▶ To study the sociohistorical correlates of these patterns of change.

Points of departure:
▶ Gender systems are very stable (Nichols 1992); they tend to “cluster in adjacent or nearby languages” (Nichols 2003: 300-303).
▶ Contact-induced loss and emergence of grammatical gender presuppose intensive bilingualism and heavy borrowing (Thomason 2001: 71).

Approach:
▶ Diachronic/dynamic typology (Greenberg 1978; Croft 2003): the study of pathways of change between language types/structures.
Domain of analysis

- **Gender agreement patterns**
  How the marking of grammatical gender on modifiers, predicates, pronouns changes over time and under the pressure of language contact.

Why?

- Inflectional morphology = morphological complexity
- Morphological complexity is sensitive to language contact dynamics (Lupyan & Dale 2010; Bentz et al. 2015).
Method

- Convenience sample of 15 sets of closely related languages (36 lngs in total), each representing:
  - Reduction/loss/expansion/emergence of gender agreement
  - A diverse range of sociohistorical profiles:
    e.g., standard/prestige languages vs. minority languages;
    high-contact varieties vs. low-contact varieties.

- Data collected through a questionnaire and descriptive resources.
Two paths of loss/reduction

1. Morphophonological erosion of agreement morphology
2. Redistribution of agreement patterns
Reduction/loss by morphophonologica l erosion

Standard Swedish (Indo-European, Germanic)

- Adnominal gender agreement: Common vs. Neuter Gender
  - *en person* ‘a person’
  - *ett hus* ‘a house’

<table>
<thead>
<tr>
<th>Table 2: Personal Pronouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hum. and Higher Anim.</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>Inanim.</td>
</tr>
</tbody>
</table>

- Comparative evidence from other Swedish dialects
  - **Elfdalian Sw.**: tripartite gender system maintained throughout the agreement system (Åkerberg 2012)
  - **Karleby Sw.**: complete gender loss except for definite article, personal and demonstrative pronouns (Huldén 1972; Hultman 1894).
Reduction/Loss by redistribution

(1) Axó Cappadocian (Indo-European, Greek; Karatsareas 2014: 79-80)

\( t \text{ spitçú} \ ta \ ndix(u)s \ xtizména \)
\[ \text{DEF.SG.GEN} \text{ house.SG.GN} \text{ DEF.PL wall.PL} \text{ built.PL}. \]

‘The walls of the house (are) built.’

(2) Modern Standard Greek (Indo-European, Greek; Karatsareas 2014: 79-80)

\( i \text{ tíçi} \ \text{ ine} \ xtixméni \)
\[ \text{DEF.M.PL wall.M.PL} \text{ be.PRS.3PL built.M.PL}. \]

‘the walls are built’.

- Comparative evidence from other Asia Minor Greek dialects:
  - **Pontic Greek**: the expansion of neuter agreement is *semantically* and *syntactically* constrained
  (inanimate nouns, agreement targets non-adjacent to nouns).
Emergent gender agreement patterns

The diachrony of many gender systems “can at best be reconstructed, but not directly observed” (Luraghi 2011: 435).

- Focus of the project: young, and grammatically non-pervasive gender systems.
  1. Resulting from light nouns, e.g., ‘man’, ‘woman’, grammaticalizing as anaphoric devices (Wälchli accepted)
  2. Resulting from borrowing of nouns and agreeing adnominal modifiers.
Emergent gender agreement patterns

The diachrony of many gender systems “can at best be reconstructed, but not directly observed” (Luraghi 2011: 435).

- Focus of the project: young, and grammatically non-pervasive gender systems. Two types:
  1. Resulting from light nouns, e.g., ‘man’, ‘woman’, grammaticalizing as anaphoric devices (Wälchli accepted)
  2. Resulting from borrowing of nouns and agreeing adnominal modifiers.
Borrowed gender agreement

Languages

- Chamorro (Austronesian)
  Contact language: Spanish
- Lekeitio Basque (Basque)
  Contact language: Spanish
- Schumcho, Jangshung (Bodic, Thebor)
  Contact language: Northern India Indo-European languages

Shared characteristics

- Gender agreement patterns passed through borrowing of inflected forms.
- Gender agreement targets are a closed class of property words.
- Gender agreement patterns are always semantic (natural gender distinctions).
Borrowed gender agreement

Chamorro (Austronesian, Marian Islands; Huber 2011: 67)

(3) Chamorro Feminine Gender (Stolz 2012: 123)

Ma-nobena-na-ye  i  mi-milagros-a  na
PASS-novena-RED-REF  DEF  abound-miraculous-F  LINK
Bithen.
Virgin

‘A novena is being conducted for the abundantly miraculous Virgin.’

(4) Chamorro Non-Feminine Gender (Stolz 2012: 125)

desde  antitites  na  tiempo  esta  gof  bunit-u  na
since  RED:before  LINK  time  already  very  nice-NF  LINK
siuda  i  ya  Hagåtña.
town  DEF  TN  Hagåtña

‘A very long time ago, Hagåtña was a very pretty town already.’
Attributive modifiers and personal pronouns have a special status in the unfolding of diachronic change in the domain of gender marking.

Under gender reduction/loss,

- The direction of change differs depending on the process involved – morphophonological erosion vs. redistribution
- Two distinct functional principles can motivate this directionality:
  - the syntactic cohesion between agreement targets and controller nouns
  - the sensitivity of different agreement targets to the semantic properties of nouns and discourse referents.
Within Eurasia, patterns of change cluster around language-family edges

<table>
<thead>
<tr>
<th>Languages</th>
<th>Family</th>
<th>Contact family</th>
<th>Observed pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cappadocian Greek</td>
<td>Greek</td>
<td>Turkic</td>
<td>Loss</td>
</tr>
<tr>
<td>Tamian Latvian</td>
<td>Balto-Slavic</td>
<td>Finnic</td>
<td>Loss</td>
</tr>
<tr>
<td>Aghul, Udi</td>
<td>Lezgic</td>
<td>Turkic</td>
<td>Loss</td>
</tr>
<tr>
<td>Karleby Swedish</td>
<td>North Germanic</td>
<td>Finnic</td>
<td>Near-loss</td>
</tr>
<tr>
<td>Kelasi, Kaftej</td>
<td>Northwestern Iranian</td>
<td>Turkic</td>
<td>Loss and expansion</td>
</tr>
<tr>
<td>Lekeitio Basque</td>
<td>Basque</td>
<td>Ibero-Romance</td>
<td>Emergence</td>
</tr>
<tr>
<td>Shumcho, Jangshung</td>
<td>Thebor</td>
<td>Indo-Aryan</td>
<td>Emergence</td>
</tr>
</tbody>
</table>

- Outlier languages within a family are neighbor with each other.
- This is in alignment with Nichols’ (2003) observation whereby grammatical gender is a cluster phenomenon.
Asymmetries in the relationship between languages in contact (may) explain the direction of change

- Contact-induced loss and emergence of gender agreement morphology presuppose prolonged contact and extensive bilingualism.

- The direction of change is to some extent predicted by the prestige dynamics between the languages in contact

<table>
<thead>
<tr>
<th>Languages</th>
<th>Change</th>
<th>Dominant contact lng</th>
<th>GG in the dominant lng</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aghul, Udi (Lezgic)</td>
<td>Loss</td>
<td>Azerbaijani (Turkic),</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Georgian (Kartvelian)</td>
<td></td>
</tr>
<tr>
<td>Igo (Ghana-Togo-Mountain)</td>
<td>Loss</td>
<td>Ewe (Gbe)</td>
<td>NO</td>
</tr>
<tr>
<td>Tamian Latvian (Balto-Slavic)</td>
<td>Loss</td>
<td>Livonian, Estonian (Finnic)</td>
<td>YES</td>
</tr>
<tr>
<td>Chamorro (Chamorro)</td>
<td>Emergence</td>
<td>Spanish (Romance)</td>
<td>YES</td>
</tr>
<tr>
<td>Lekeitio Basque (Basque)</td>
<td>Emergence</td>
<td>Spanish (Romance)</td>
<td>YES</td>
</tr>
</tbody>
</table>
To sum up

This project has contributed to highlight:

▶ types of changes in gender agreement systems and possible directionalities in the spread of these changes
▶ a number of sociohistorical variables that are relevant to the understanding of the evolution of gender agreement systems.

Limitations:

▶ Only a limited number of languages per family.
▶ Too little data for some of the languages in the sample.
Next step

- Using the results of the study as a starting point for further hypothesis testing on larger data sets (one family in detail), and with the support of quantitative methods.
Correlates of restructuring in Bantu gender systems (with Annemarie Verkerk, MPI – Jena)

- Studying the diversity of the gender systems of the Bantu languages.
- Testing the models of language change that account best for within-family variation in this domain of grammar.
- Investigating socio-historical correlates of the distribution of this variation.
(5) Gender marking in Chichewa (Kiso 2012: 18)

chi-nkhanira cha-\text{ch}i-kazi  
CL7-scorpion ASS-CL7-female  
chi-ku-dzi-kanda  
CL7.SBJ-PRES-REFL-scratch

“The female scorpion is scratching itself”.

(6) Gender marking in Kinshasa Lingala (Meeuwis 2013: 30)

a. Mw-ana a-ko-kweya  
CL1-child 3SG.ANIM-FUT-fall

‘The child will fall.’

b. Ndako e-ko-kweya  
CL9.book 3SG.INAN-FUT-fall

‘The house will fall.’
The Bantu languages and their gender systems

(7) Gender marking in Chichewa (Kiso 2012: 18)

\[ \text{chi-nkhanira cha-chi-kazi} \]
CL7-scorpion ASS-CL7-female

\[ \text{chi-ku-dzi-kanda} \]
CL7.SBJ-PRES-REFL-scratch

“The female scorpion is scratching itself”.

(8) Gender marking in Kinshasa Lingala (Meeuwis 2013: 30)

a. \[ \text{Mw-ana a-ko-kweya} \]
CL1-child 3SG.ANIM-FUT-fall
‘The child will fall.’

b. \[ \text{Ndako e-ko-kweya} \]
CL9.book 3SG.INAN-FUT-fall
‘The house will fall.’

Questions

▶ How do we go from the Chichewa type to the Kinshasa Lingala type?
▶ Why does this happen?
The evolution of Bantu gender marking systems

Questions

1. Which word classes carry gender marking besides nouns (e.g., pronouns, verbs, adjectives)?
2. Are animacy-based distinctions part of the gender marking system?

Quantitative data analysis

The coding will be mapped on the Bantu phylogenetic tree (Grollemund et al. 2015) to estimate transition probabilities between attested systems using Phylogenetic Comparative Methods.

Hypotheses

1. Animacy-based distinctions encroach the gender marking system starting from anaphoric pronouns and gender markers on verbs.
2. Marking on nouns is more stable than marking on other word classes.
(Wald 1975; Di Garbo & Miestamo accepted)
Sociohistorical correlates

- Variables we plan to work with:
  - Population data (both L1 and L2)
  - Presence/absence of gender systems in neighboring languages

Hypotheses

1. Large populations with high proportions of L2 users and intense language contact predict reduction and/or loss of gender marking.
2. Small populations with low proportions of L2 users and intense language contact predict retention of gender marking (group identity marking) or its reduction/loss (shift-induced interference).
3. Geographic proximity between related and unrelated languages predicts convergence in the domain of gender marking.
What we’ve done so far

▷ Defined the coding procedure
▷ Collected data about the gender system of 130+ Bantu languages.
Maho’s (1999) classification of Bantu gender systems

<table>
<thead>
<tr>
<th>Nouns</th>
<th>1= Tr.</th>
<th>2= Tr. + An.</th>
<th>2^i = Tr+Pl</th>
<th>3 = An. +Sg/Pl</th>
<th>4 = Sg/Pl</th>
<th>5= None</th>
</tr>
</thead>
</table>

Elsewhere
A = Tr.
B = Tr. + An.
C = An. + Sg/Pl
D = Sg/Pl
E = None
### Maho’s (1999) typology of Bantu gender systems

<table>
<thead>
<tr>
<th>Nouns</th>
<th>1 = Tr.</th>
<th>2 = Tr. + An.</th>
<th>2(^i) = Tr+Pl</th>
<th>3 = An. + Sg/Pl</th>
<th>4 = Sg/Pl</th>
<th>5 = None</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Elsewhere</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A = Tr.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B = Tr. + An.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C = An. + Sg/Pl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D = Sg/Pl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E = None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Zulu</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Swahili</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lunda</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>K. Lingala</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Amba, Bera</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pande</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Homa</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Kongo</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Kituba</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Komo</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The languages of the sample based on Maho's types
Some observations

- The gender systems of several Bantu languages show a bias towards the overt expression of animacy distinctions.
- The spread of this feature within the family is NOT a unitary process.
- Multiple developments must be posited in different subareas of the Bantu speaking world.
The northern Bantu borderlands

Radically restructured gender systems in the Bantu northern borderlands

Legend
- Type 1: N_tr; AG_an
- Type 2: N_an; AG_an
- Type 3: N_an; AG_sg/pl
- Type 4: N_rel; AG_none
- No radical restructuring

Two possible scenarios:

1. Substratum interference from pre-Bantu populations shifting to Bantu languages, possibly including Pygmies.
2. Continued contact between Bantu and non-Bantu languages in the area.
Thank you very much!

Thanks are also due to:

Anna Ahlströms och Ellen Terserus Stiftelse
References


Di Garbo, Francesca. under review. The complexity of gender and language ecology, .


