



# On the problem of categorization in linguistics

Rik De Busser

VNU-ULIS, Friday 25 October 2019



政大  
NATIONAL CHENGCHI UNIVERSITY

# Introduction

- What is categorization?
  - Why do we need it?
- Problems with categorization in linguistics
- Word classes
  - Historical background
  - Case studies
- Discussion
  - English revisited
  - Vietnamese

# Categorization

- Classifying object, events, ... into categories
- General cognitive process (so not specific to language)
- Extremely widespread function that influences many aspects of our daily lives
- Basic survival skill to humans and other living beings
- Crucial in the normal operation of every language

Feline

Lynx

Cat



# Conceptual categories

- What is it that we are actually categorizing?
  - What are these conceptual categories?
  - What are we categorizing with these categories?
- Categorizing cognitive concepts
- Two levels of analysis:
  - Category membership of concepts
    - E.g. What is the semantic relationship between different types of birds ('robin', 'lark', 'ostrich') and the general concept 'bird'
  - Assignment of concepts to entities and events
    - E.g. In the real world, which animals can be called 'bird'

# Theories of categorization

- Classical models
- Exemplar-based models
- Prototype models
  
- Probabilistic models

# The classical model

1. Categories are defined by a set of essential features
  - Peripheral features are not important in defining a category
2. Category membership is determined by the presence of these essential features
  - Essential features hold a category together, so all member should have them
3. All category members have equal status
  - If all members have the same essential features, they are all equally good representatives of their category
4. Category boundaries are sharp and rigid
  - If a concept has all essential features, it is a member of the category; if not, it cannot be a member

# The classical model

- A modern version: componential analysis
- A way of formalizing semantic analysis based on essential features

MAMMAL  $\left[ \begin{array}{l} +hair \\ -feathers \\ -egg\_laying \\ -wings \\ +warm\_blooded \end{array} \right]$

BIRD  $\left[ \begin{array}{l} -hair \\ +feathers \\ +egg\_laying \\ +wings \\ -warm\_blooded \end{array} \right]$

# Exemplar-based models

- Categories are defined by a number of typical examples (exemplars)
  - These exemplars will have a number of salient features
  - ... but there is no expectation that there is an overlap of features
- Category membership is determined by the similarity of the member and exemplar memories
  - No rules of comparison
  - Comparison of essential and non-essential features
  - Distance function determines how close the member is to its exemplar
- Certain members are more central than others
- Categories are not fixed and have no clear boundaries

## Exemplars



{  
+feathers  
+beak  
+bipedal  
+colorful  
+blue  
sound: week week  
+flies  
size: 12 – 23cm  
}



{  
+feathers  
+beak  
+bipedal  
+colorful  
+orange  
sound: tsip tsip  
+flies  
size: 6 – 9cm  
}



???

# Exemplars



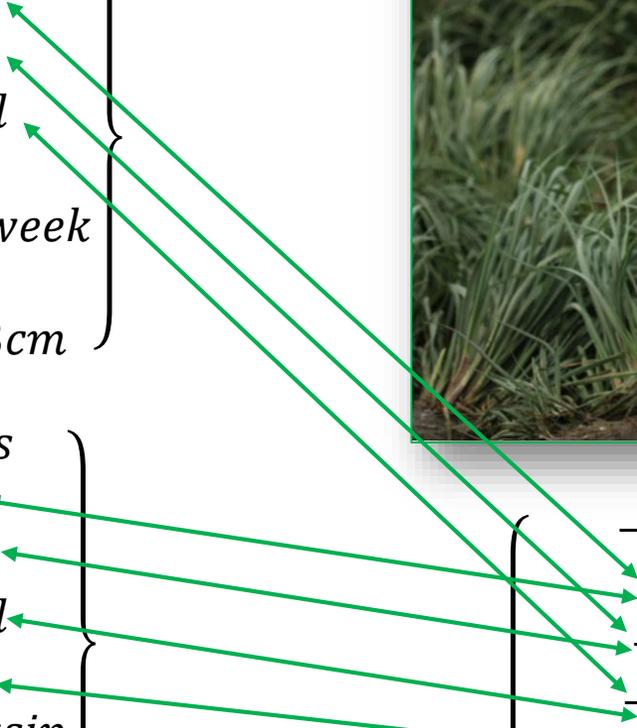
{  
+feathers  
+beak  
+bipedal  
+colorful  
+blue  
sound: week week  
+flies  
size: 12 – 23cm  
}



{  
+feathers  
+beak  
+bipedal  
+colorful  
+orange  
sound: tsip tsip  
+flies  
size: 6 – 9cm  
}

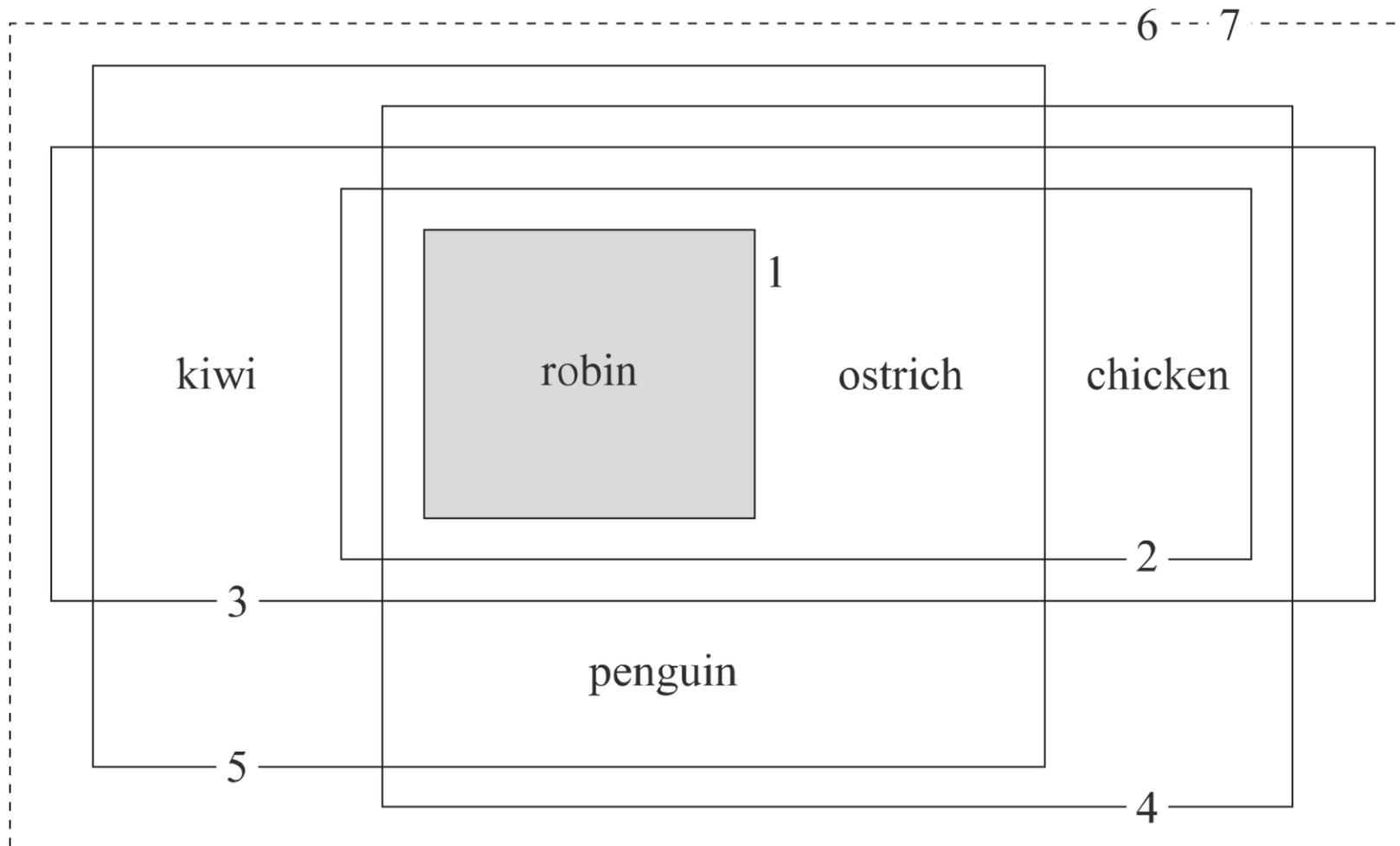


{  
-feathers  
+beak  
+bipedal  
+colorful  
+orange  
+yellow  
sound: gwaa gwaa  
-flies  
size: 45 – 60cm  
}



# Prototype models

1. Categories are defined relative to prototypes
  - Prototypes can be members of a category with typical features or they can be abstract feature bundles
  - Most or even all members of a class do not have all typical features
2. Category membership is defined by family resemblance
  - Centrality to a class is determined by how similar a member is to its prototype
3. Certain members of a category are more central than others
  - Members in a category are organized in a “radial set of clustered and overlapping meanings” (Geeraerts 2006)
4. Categories can be flexible and have fuzzy boundaries, but they can also have sharp and fixed boundaries



1 being able to fly

4 having wings

7 having a beak or bill

2 having feathers

5 not domesticated

3 being S-shaped

6 being born from eggs

# From theory to practice

- Categorization is important in language
  - Semantic categories
    - E.g. humans, animate (living) objects, animals, birds, food, ...
  - Syntactic categories
    - E.g. nouns, verbs, personal pronouns, tenses, ...
  - Socio-cultural categories
    - E.g. people of authority, parents vs. children, older vs. younger people, ...
- We can use different models to categorize concepts
- We can categorize concepts at different levels of generality
- How does this all work together?

# A case study: nouns and verbs

- Word classes are at the basis of any linguistic analysis
- They are often not well-defined
- Do all languages have nouns and verbs?
- If so, how do we define them?
- What about adjectives and adverbs?
- Where do these word classes come from?

# A history of word classes

- Word classes as we know them today did not always exist
- They slowly developed in the Western philosophical tradition
- ‘Modern’ word classes only came into existence around the 4<sup>th</sup> century
- Word classes ...
  - are no natural categories
  - were developed in a Western framework
  - were used more or less unchanged for 16 centuries

# Aelius Donatus

- Mid 4<sup>th</sup> century
- Teacher of St Jerome
  - Church Father
  - North Eastern Italy
  - Translator of the Vulgate, the Latin translation of the Bible
- Nothing else known about his life

Saint Jerome  
Albrecht Dürer, ca. 1495  
National Gallery London



# Aelius Donatus

- Donatus Orthographicus
- Works
  - Ars grammatica ‘The art of grammar’
    - Ars minor
    - Ars maior
  - Commentarii Vergiliani ‘Commentaries on the life of Virgil’



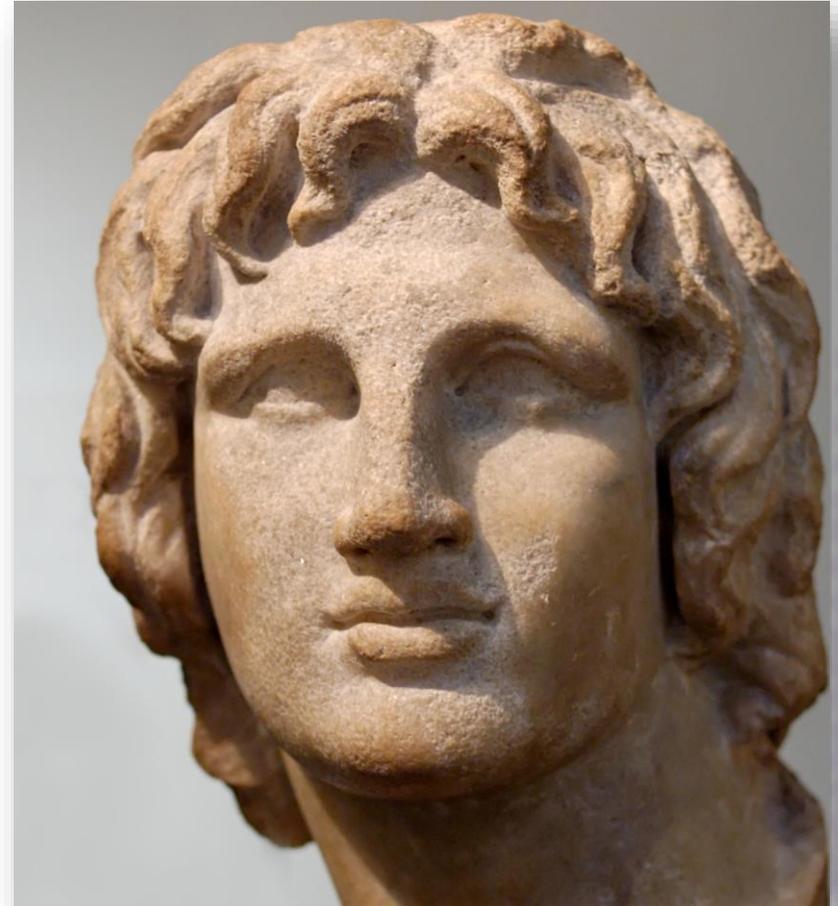
(Nuremberg Chronicle,  
1493)

# Background

- Pedagogical grammars (see Harris & Taylor 1997, Ch. 4)
- Study of Greek and later Latin
- European culture
- From late Antiquity to early Middle Ages

# Hellinistic Period

- 323 - 31 BC
- Alexander the Great's empire
- Creation of a standard language (based on Attic Greek)
  - ... And an associated written tradition
  - Writing as part of imperial administration
  - Greek language learning as a way to incorporate conquered people
- Need for education scribes and scholars in the Greek language



# Hellinistic Period

- Shift:

Spoken language



Written language



Standardization

# From late Antiquity onwards

- 3<sup>rd</sup> – 8<sup>th</sup> century
- Description of grammar as a set of systematic rules
- *Techne grammatike* (τέχνη γραμματική)
  - Ascribed to Dionysius Thrax (Διονύσιος ὁ Θρᾶξ) (ca. 170 - 90 BC)
  - Introduction of grammar as a separate study subject for scholars
  - Eight parts of speech
  - Influence on later scholars, including Donatus

# Περὶ γραμματικῆς / About grammar

γραμματικὴ ἐστὶν ἐμπειρία τῶν παρὰ ποιηταῖς τε καὶ συγγραφεῦσιν ὡς ἐπὶ τὸ πολὺ λεγομένων.

μέρη δὲ αὐτῆς ἐστὶν ἕξ·  
πρῶτον ἀνάγνωσις ἐντριβῆς κατὰ προσωιδίαν,  
δεύτερον ἐξήγησις κατὰ τοὺς ἐνυπάρχοντας ποιητικοὺς τρόπους,  
τρίτον γλωσσῶν τε καὶ ἱστοριῶν πρόχειρος ἀπόδοσις,  
τέταρτον ἐτυμολογίας εὔρεσις,  
πέμπτον ἀναλογίας ἐκλογισμὸς,  
ἕκτον κρίσις ποιημάτων, ὃ δὲ κάλλιστόν ἐστι πάντων τῶν ἐν τῇ τέχνῃ.

(Technē grammatikē A.α)

Grammar is the practical study of the usage of poets and prose writers.

It has six parts: first, correct pronunciation of the text, paying attention to the diacritics; second, explanation of the poetic tropes encountered; third, elucidation of any difficult words or allusions; fourth, the tracing of etymologies; fifth, the establishment of analogical regularities; sixth, literary appreciation, which is the finest part of all.

(Harris & Taylor 1993: 48)

# Early Middle Ages

- Before 11<sup>th</sup> century
- Interpretation of Latin and Greek texts
  - Classical Greek and Latin scholars
  - Christian texts
- In a scholastic tradition
  - Christian scholars
  - Translation and interpretation of Classical texts in a Christian context
  - Translation and correct interpretation of the Bible and other religious texts
  - Religious orthodoxy
- Basis for scholarship in the West until today

# Reading Donatus today

- Late antiquity
- What was the impact of this scholarship on linguistic tradition?
- Similarities and differences with today's methods of analysis?
- Limitations?

# DE NOMINE

nomen quid est? pars orationis cum  
casu corpus aut rem proprie  
communiterue significans.

nomini quot accidunt? sex.

quae? qualitas comparatio genus  
numerus figura casus.

# Of the noun

What is a noun? A part of speech which  
with the case a person or a thing  
specifically or generally.

How many accidents (attributes) does it  
have? Six.

Which ones? Quality, comparison,  
gender, number, form and case.

# DE NOMINE

qualitas nominum in quo est? bipertita est: aut enim unius nomen est et proprium dicitur, aut multorum appellatiuum.

comparationis gradus quot sunt? tres.

qui? positius, ut doctus, comparatiuus, ut doctior, superlatiuus, ut doctissimus.

# Of the noun

Wherein lies the quality of a noun? It is twofold: either it is a single name, and it is called a proper noun, or it refers to many [referents].

How many grades of comparison are there? Three.

Which ones? The positive, as in 'learned'; the comparative, as in 'more learned'; and the superlative, as in 'most learned'.

# DE NOMINE

quae nomina comparantur? appellatiua dumtaxat qualitatem aut quantitatem significantia.

comparatiuus gradus cui casui seruit? ablatiuo sine praepositione: dicimus enim doctior illo.

superlatiuus cui? genetiuo tantum plurali: dicimus enim doctissimus poetarum.

# Of the noun

What kind of nouns are compared? Only common nouns signifying quality or quantity.

What case is the comparative degree used with? The ablative without a preposition; for we say 'more learned than he'.

What case with the superlative? Only the genitive plural: for we say 'most learned of poets'.

# Summary

- Nouns according to Donatus
  - A part of speech
  - Referring to persons or things
  - Having the following syntactic properties: quality, comparison, gender, number, form and case
  - Distinction between common and proper nouns
  - Nouns referring to qualities (our adjectives) have grades of comparison
- Pretty much our modern noun + adjective
- Adjectives in Latin were noun-like

# Summary

- The influence of Aelius Donatus on the language studies and teaching was immense
  - Same parts-of-speech still used today
  - Greek and Latin were considered the basis for analysis
    - Rich morphology
    - Cases for nouns, complex paradigms for verbs

FIRST (OR  $\bar{A}$ -) CONJUGATION.

101. Active Voice. — Amō, *I love.*

PRINCIPAL PARTS.

PRæs. IND.	PRæs. INF.	PERF. IND.	PERF. PASS. PARTIC.
amō	amāre	amāvī	amātus

INDICATIVE MOOD.

PRESENT TENSE.

SINGULAR.	PLURAL.
amō, <i>I love,</i> amās, <i>you love,</i> amat, <i>he loves;</i>	amāmus, <i>we love,</i> amātis, <i>you love,</i> amant, <i>they love.</i>

IMPERFECT.

amābam, <i>I was loving,</i> amābās, <i>you were loving,</i> amābat, <i>he was loving;</i>	amābāmus, <i>we were loving,</i> amābātis, <i>you were loving,</i> amābant, <i>they were loving.</i>
--	--

FUTURE.

amābō, <i>I shall love,</i> amābis, <i>you will love,</i> amābit, <i>he will love;</i>	amābīmus, <i>we shall love,</i> amābitis, <i>you will love,</i> amābunt, <i>they will love.</i>
--	---

PERFECT.

amāvī, <i>I have loved, I loved,</i> amāvistī, <i>you have loved, you loved,</i> amāvit, <i>he has loved, he loved;</i>	amāvīmus, <i>we have loved, we loved,</i> amāvistis, <i>you have loved, you loved,</i> amāverunt, -ēre, <i>they have loved, they loved.</i>
---	---

PLUPERFECT.

amāveram, <i>I had loved,</i> amāverās, <i>you had loved,</i> amāverat, <i>he had loved;</i>	amāverāmus, <i>we had loved,</i> amāverātis, <i>you had loved,</i> amāverant, <i>they had loved.</i>
--	--

FUTURE PERFECT.

amāverō, <i>I shall have loved,</i> amāveris, <i>you will have loved,</i> amāverit, <i>he will have loved;</i>	amāverīmus, <i>we shall have loved,</i> amāveritis, <i>you will have loved,</i> amāverint, <i>they will have loved.</i>
--	---

SUBJUNCTIVE.

PRESENT.

SINGULAR.	PLURAL.
amem, <i>may I love,</i> amēs, <i>may you love,</i> amet, <i>let him love;</i>	amēmus, <i>let us love,</i> amētis, <i>may you love,</i> ament, <i>let them love.</i>

IMPERFECT.

amārem, <i>I should love,</i> amārēs, <i>you would love,</i> amāret, <i>he would love;</i>	amārēmus, <i>we should love,</i> amārētis, <i>you would love,</i> amārent, <i>they would love.</i>
--	--

PERFECT.

amāverim, <i>I may have loved,</i> amāveris, <i>you may have loved,</i> amāverit, <i>he may have loved;</i>	amāverīmus, <i>we may have loved,</i> amāveritis, <i>you may have loved,</i> amāverint, <i>they may have loved.</i>
---	---

PLUPERFECT.

amāvīsem, <i>I should have loved,</i> amāvissēs, <i>you would have loved,</i> amāvisset, <i>he would have loved;</i>	amāvissēmus, <i>we should have loved,</i> amāvissētis, <i>you would have loved,</i> amāvissent, <i>they would have loved.</i>
--	---

IMPERATIVE.

Pres. amā, <i>love thou;</i>	amāte, <i>love ye.</i>
Fut. amātō, <i>thou shall love,</i> amātō, <i>he shall love;</i>	amātōte, <i>ye shall love,</i> amantō, <i>they shall love.</i>

INFINITIVE.

Pres. amāre, *to love.*  
Perf. amāvīsse, *to have loved.*  
Fut. amātūrus esse, *to be about to love.*

PARTICIPLE.

Pres. amāns,<sup>1</sup> *loving.*  
(Gen. amantis.)  
Fut. amātūrus, *about to love.*

GERUND.

Gen. amandī, *of loving,*  
Dat. amandō, *for loving,*  
Acc. amandum, *loving,*  
Abl. amandō, *by loving.*

SUPINE.

Acc. amātum, *to love,*  
Abl. amātū, *to love, be loved.*

# Summary

- The influence of Aelius Donatus on the language studies and teaching was immense
  - Same parts-of-speech still used today
  - Greek and Latin were considered the basis for analysis
    - Rich morphology
    - Cases for nouns, complex paradigms for verbs
  - Influence on
    - Definition of word classes  $\Rightarrow$  reliance on morphology
    - Perception of other languages

# Perception of other languages

“We must not forget that there are languages which have remained in that germinal state, and in which there is to the present day no *outward* distinction between a root and a word. In **Chinese**, for instance, ...”

Max Mueller, 1864,  
*Lectures on the Science of Language, Second Series*  
(quoted in Harris & Talbot 1997, p. 54-55)

⇒ Languages without morphology are ‘germinal’ or ‘primitive’

# Beyond the Western tradition

- How do these categories work for non-Western languages?
- Do major word classes exist in all languages?
- Focus on nouns and verbs
- Two possible hypotheses:
  1. Word classes reflect ingrained cognitive categories that are common to all humans
  2. Word classes reflect relative categories determined by the Western tradition and its languages

# Beyond the Western tradition

- Case studies
  - Bunun
  - Mandarin Chinese
  - (Vietnamese & English)

# Categorizing word classes

- Word classes are determined language-internally
  - Distinctive properties: properties that can tell word classes apart
  - Typical properties: meaningful properties that are typically (but not necessarily always) associated with a word class
- Nature of the evidence (Evans 2000)
  - (Phonology and prosody)
  - Semantics
  - **Morphosyntax**
  - Functional information

# Categorizing word classes

- Primary distinction between nouns and verbs
  - Distinguishing between Entities (people and things) and Events (actions, states, etc.)
  - In many languages, other word classes can be defined at least partly by how much they look like nouns or verbs
    - E.g. across languages, adjectives are often noun-like or verb-like
- Typical properties used to distinguish nouns and verbs:

# Noun-verb distinction

## **Nouns**

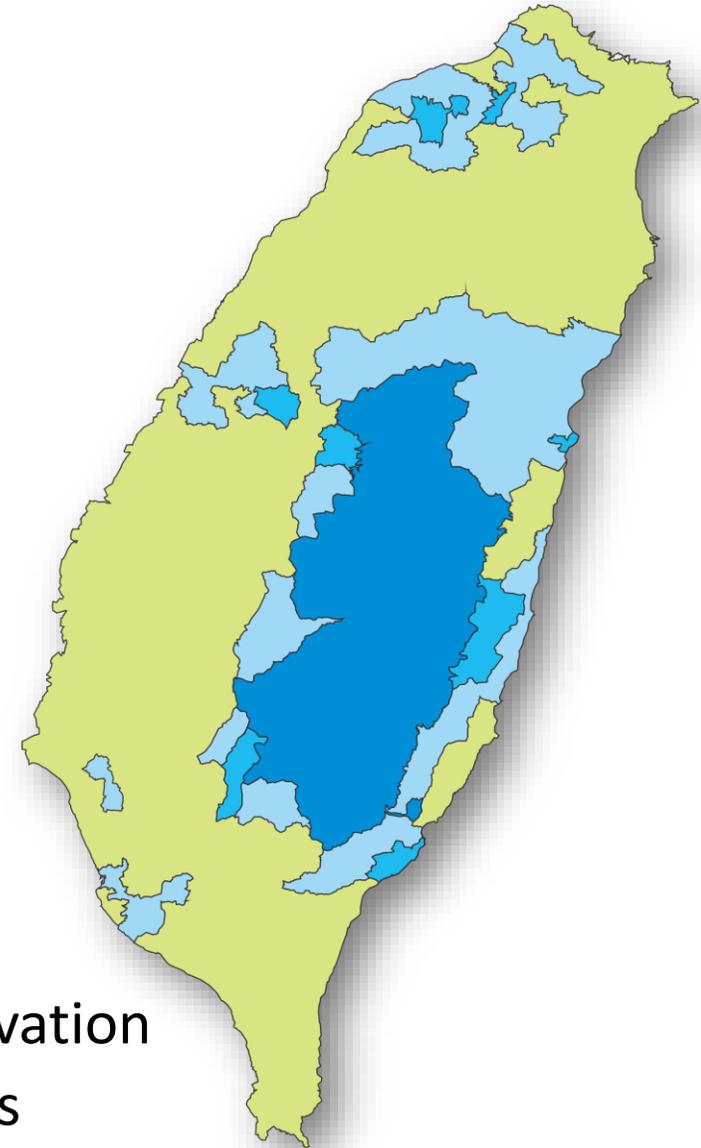
- Typically refer to persons or objects
- Typically function as arguments
- Can get definiteness marking
- Typically mark case
- Typically mark number
- Typically mark gender

## **Verbs**

- Typically refer to actions and states
- Typically function as predicates
- Typically can get voice marking
- Typically mark tense
- Typically mark aspect
- Typically mark mood

# Bunun

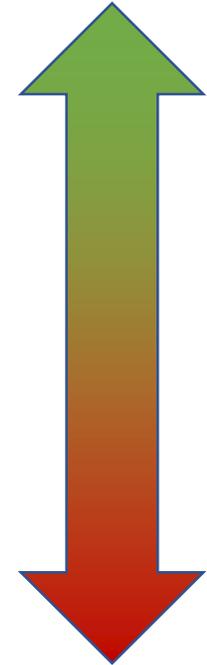
- Austronesian language, Taiwan
- Five dialects
  - Northern: Takibakha, Takituduh
  - Central: **Takbanuaz, Takivatan**
  - Southern: Isbukun
- Philippine-type voice system (De Busser 2011)
  - AV, UV, LV, ...
- Very productive morphology
  - No clear distinction between inflection and derivation
  - For instance, voice markers are also nominalizers



# Nouns and verbs in Austronesian languages

Everything is normal (but in a weird way)

- Traditional categories (Kroeger 1998)
- Non-traditional categories (Himmelman 2008)
- Nominalist hypothesis (Kaufman 2009a, 2009b)
- (Root) precategoriality (Foley 1998)
- There are no word classes (Broschart 1997; Gil 1994, 2009)



All word classes are an illusion

# Nouns and verbs in Bunun

- There appear to be nouns and verbs
  - Semantics
  - Typical use

# Bunun aspect marking: progressive *-an*

- (1) ... maupata sia lainiqaiban tu isanaŋ Sipun.  
maupa=ta                    sia            l<in>aqaiban tu            **i-san-aŋ**                    Sipun  
thus=ART.ENT.DIST    ANAPH    <PST>route            COMPL LOC-be.at-PROG Japanese  
'... our lives were like that when the Japanese were still present.'
- (2) ... daŋiʔanani tinasʔi kaku tudipʔað Sipunaŋ.  
daŋi-an-an-i                    t<in>asʔi            kaku            tudip-ʔað                    **Sipun-aŋ**  
place-LOCATION-LV-PRT    <PST>make            school that.time-ADJR    Japanese-PROG  
'... the location of the school that was built still by the Japanese.'
- (3) Lini            Pulaʔaŋ  
Lini            **Pula-aŋ**  
L.            P.-PROG  
'Lini, then still the husband of Pula'

# Bunun aspect marking: perfective *-in*

- (4) Mudanin su tama lumaqti?  
**mu-dan-in** su tama lumaq=ti  
ALL-go-PRV 2S.N father home=ART.ENT.PROX  
'Did your father already go home?'
- (5) Tinsihalin nak tian  
**tin-sihal-in** nak tian  
SUDDEN-good-PRV 1S.N belly  
'My belly is suddenly better.'
- (6) Han dan vasu tumvasu?in, ...  
han dan vasu **tum-vasu-in**  
at railroad RIDE-small.train-PRV  
'When we arrived at the railroad, and we took the train...'

# Bunun aspect marking: perfective *-in*

- (7) Maqi daŋku vali?in, mun?iti maun.  
maqi **daŋku vali-in** mun-?iti maun  
if zenith sun-PRV ALL-here eat  
'When it has become noon, come over here to eat.'

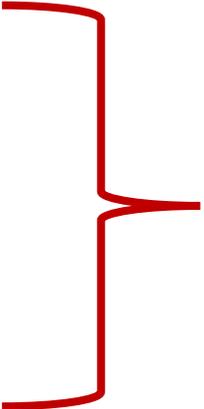
- (8) Lumaqain mal?u.  
**lumaq-in** m-al?u **N+Perfective = Predicate**  
house-PRV DYN-rest  
'She is at home doing nothing.'

# Bunun voice marking: locative voice *-an*

Daŋiʔan sak sui  
daŋi-an sak sui  
place-LV 1S.TOP.AG money  
'I take care of the money.'

Masihalan tu daŋiʔan  
ma-sihal-an tu daŋi-an  
STAT-good-LV ATTR place-LOCATION  
'It is a good place'

Tatasʔian hulus  
ta-tasʔi-an hulus  
CV-make-LOCATION clothes  
'a shop that is professionally making clothes'



**LV ≈ LOCATION?**

# Bunun voice marking: locative voice *-an*

istasʔi tulkukan

is-tasʔi

tulkuk-an

INSTR-make chicken-LOCATION

‘I use it do build a chicken coop’ (lit: ‘chicken location’)

baunan

buan-an

moon-LV

‘the moon is shining’

**LOCATION ≈ LV?**

# Nouns and verbs in Bunun

- There appear to be nouns and verbs
    - Semantics
    - Typical use
  - ... but the traditional criteria for categorizing nouns and verbs do not work
    - Both nouns and verbs can be predicates
    - Both nouns and verbs can have definiteness markers
    - Both nouns and verbs can have tense and aspect marking (but verbs more often do so)
    - Both nouns and verbs can have voice marking
    - Voice marking is also used for certain kinds of derivations
- ⇒ It is a bit of a mess

# Why would we care?

- Word classes are important in linguistics
- (And in language teaching)
- Especially nouns and verbs
- A systematic word class categorization problem in linguistics
  - Especially in lesser studied languages

# Why would we care?

- If the categories nouns and verb do not exist or cannot be established using similar criteria, a number of questions arise:
  - Are these problems an indication that Noun and Verb are not general cognitive categories?
  - Should we revise or update how we establish the basic categories in language?
  - How can we compare the grammar of languages if their word classes are not the same?
  - How can we use linguistic theories across languages?
  - How should we teach languages to our students?

# Why would we care?

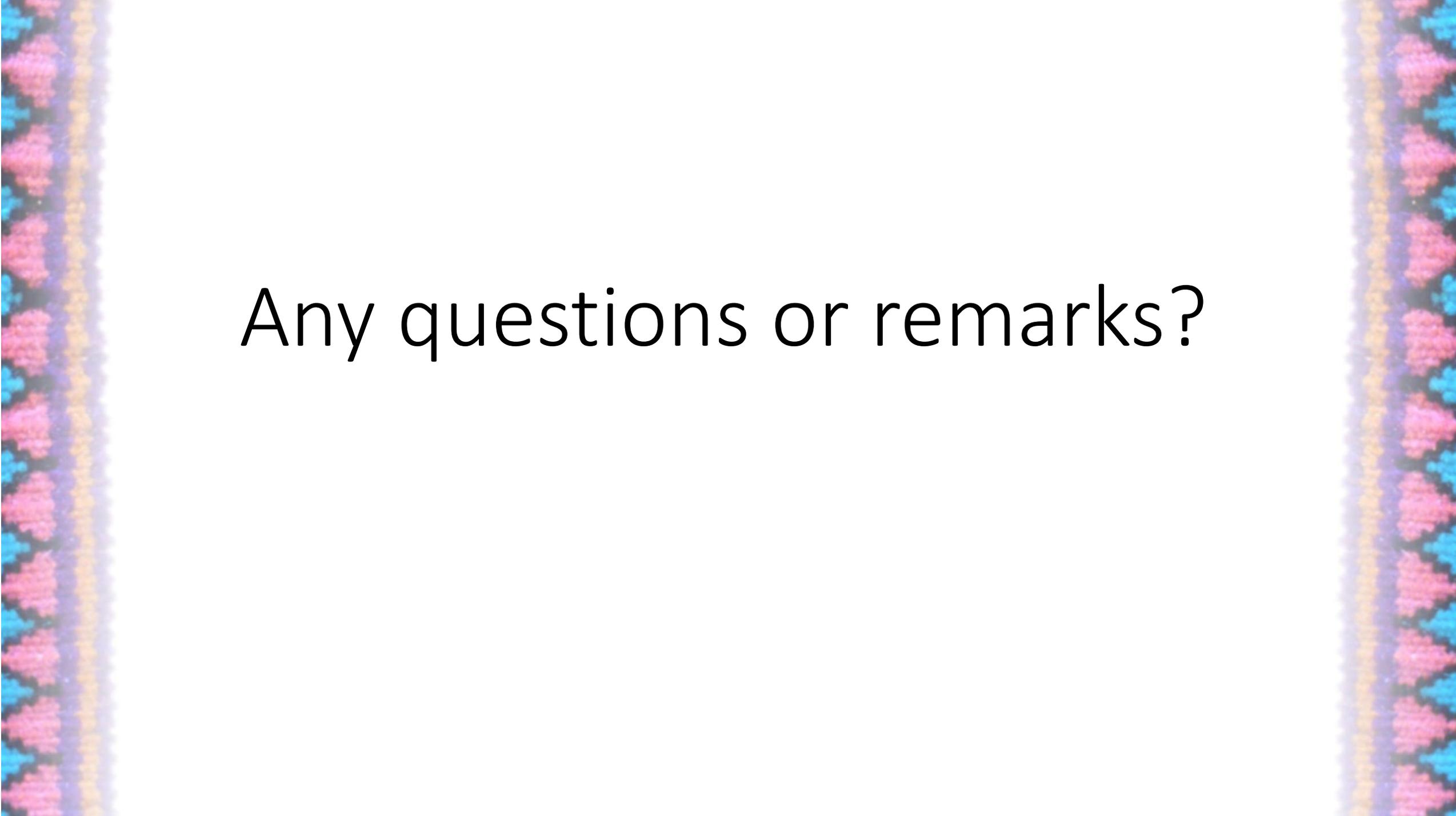
- But maybe what we discovered is not as unexpected as we thought



- Let's review some better known languages
  - How do they establish the noun-verb distinction?
  - Which problems do we encounter?



- English & Mandarin Chinese
- Vietnamese



Any questions or remarks?

# Discussion

- English noun-verb distinctions
  - Compound
  - Cry
  - Flower
  - Throw
  - Work
- We can use online corpora to analyze these examples:
  - <https://www.english-corpora.org/>

# Bibliography

Bennett, Charles E. 1908. *A Latin Grammar*. Boston: Allyn and Bacon.

Broschart, Jürgen. 1997. Why Tongan does it differently: categorial distinctions in a language without nouns and verbs. *Linguistic Typology* 1–2. 123–165.

De Busser, Rik. 2009. *Towards a Grammar of Takivatan Bunun: Selected Topics*. Melbourne: La Trobe University PhD.

De Busser, Rik. 2011. Towards an analysis of argument alignment in Takivatan Bunun. *Studies in Language* 35(3). 523–555.

Donatus, Aelius. 4th century. *De partibus orationis ars minor*. URL: <http://www.thelatinlibrary.com/don.html>

Evans, Nicholas. 2000. Word classes in the world's languages. In Geert Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung*, vol. 1, 708–732. Berlin: Walter de Gruyter.

# Bibliography

- Foley, William Auguste. 1998. Symmetrical voice systems and precategoriality in Philippine languages. Brisbane: University of Queensland.
- Geeraerts, Dirk. 2006. Prospects and problems of prototype theory. In Dirk Geeraerts (ed.), *Cognitive Linguistics: Basic Readings*, 141–165. Berlin: Mouton de Gruyter.
- Gil, David. 1994. The structure of Riau Indonesian. *Nordic Journal of Linguistics* 17(2). 179–200.
- Gil, David. 2009. Austronesian Nominalism and the Thinginess Illusion. *Theoretical Linguistics* 35(1). 95–114.
- Harris, Roy & Talbot J. Taylor. 1997. *Landmarks in Linguistic Thought I: The Western Tradition from Socrates to Saussure. Second Edition*. London: Routledge.

# Bibliography

- Himmelmann, Nikolaus P. 2009. Notes on Tagalog nominalism. *Theoretical Linguistics* 35(1). 115–123.
- Kaufman, Daniel. 2009a. Austronesian typology and the nominalist hypothesis. In K. Alexander Adelaar & Andrew K. Pawley (eds.), *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust*, 197–226. Canberra: Pacific Linguistics.
- Kaufman, Daniel. 2009b. Austronesian nominalism and its consequences: A Tagalog case study. *Theoretical Linguistics* 35(1). 1–49.
- Kövecses, Zoltán. 2006. *Language, Mind, and Culture: A Practical Introduction*. New York: Oxford University Press.
- Thrax, Dionysios. 1874. *The Grammar of Dionysios Thrax*. (Trans.) Thomas Davidson. St. Louis, MO: R. P. Studley.