Diversity and vagueness in language

PRELIMINARIES
This talk deals with two presuppositions on my part: what (I think) non-linguists believe is true for all natural human languages:
1. THEY ARE ALL RATHER SIMILAR (LIKE DANISH, ENGLISH, ETC.);
2. THEY ARE HOPELESSLY VAGUE AND UNDERSPECIFIED.

It will be shown that (within certain limits) HUMAN LANGUAGES
- ARE EXTREMELY VARIED IN ALL DOMAINS: SYNTACTICALLY, MORPHOLOGICALLY,
  SEMANTICALLY, PRAGMATICALLY, …; and
- CAN BE EVEN MORE (MUCH MORE) UNDERSPECIFIED THAN DANISH OR
  ENGLISH.

I will also argue, however, that there is unity in all this diversity, and that underspecification is often necessary for successful communication.

- How many languages?
Ca. 6000 (includes many ‘dialects’).
Language or dialect? Weinreich: “A language is a dialect with an army”
Some predict that by the end of this century approximately 50% of the languages that are currently spoken will be extinct (Krauss 1992: 5)

- Multilingualism
Most people are multilingual. The Brazilian Vaupés area of Western Amazonia, where several Tukanoan languages and one Arawakan language (Tariana) co-exist, is famous for its linguistic exogamic rules: marriages are possible only if the spouses speak different languages (one’s language is defined as one’s father’s language). [Sorenson 1967; Meira 2005: 1698]

- Historic relations - how many Language Families: 1? 30? 120? …
☞ “Lumpers” (few families) vs. “Splitters” (many families)
☞ SEE MAPS below (from Ruhlen 1987 = “lumper”).
1. DIVERSITY: HOW LANGUAGES DIFFER

Mirror patterns in Japanese and English (Smith 1978: 78)

(1) Toîkyoî eki kara denîsha de ichijikani kurai nishi e Tokyo station from tram by one hour about west toward
itta tokoro ni Kamurako to iu machi ga arîmasu go place at Kamakura “Quote” called town OBJ exists

(2) There is a town called Kamakura at a place (you can reach)

(3) SOV < SVO < VSO < VOS < OVS < OSV

Japanese:

(4) Subject Object Verb
Taroo ga tegami o kaita
‘Taroo wrote a letter’

Danish:

(5) Subject Verb Object
Vibeke skrev et brev

Welsh:

(6) Verb Subject Object
Lladdodd y ddraig y dyn killed the dragon the man
‘The dragon killed the man’

Malagasy:

(7) Verb Object Subject
Nahita ny vehivavy ny mpianatra saw the woman the student
‘The student saw the woman’

Hixkaryana:

(8) Object Verb Subject
Toto yonove kamara man he-ate-him jaguar
‘The jaguar ate the man’

Syntax (sentence structure): Subject - Verb - Object can be ordered in six different ways:
SOV (man house build), SVO (man build house), VSO (build man house), VOS (build house man), OVS (house build man), and OSV (house man build).

Relative frequency of basic word orders (preference S before O):
44.78% 41.79% 9.20% 2.99% 1.24% ? (n=420)
**Morphology (word structure):** separate words (isolating) or all-in-one (polysynthetic)

**Beijing Chinese, China** (Chrysant 1987: 293)

(9) \[\text{Wō māǐ júzǐ chī} \]
    \begin{align*}
    & \text{I buy} \\
    & \text{orange eat} \\
    & \text{‘I bought some oranges to eat’}
    \end{align*}

**Swahili, Kenya** (Chrysant 1987: 293)

(10) \[\text{mimi ni-na-ku-penda wewe} \quad \text{[mini ninakupenda wewe]} \]
    \begin{align*}
    & \text{me I-PRESENT-you-love you} \\
    & \text{‘I love you’}
    \end{align*}

**Tiwi, Australia** (Chrysant 1987: 293)

(11) \[\text{ngi-rru-ntihing-apu-kani} \quad \text{[ngirruunthingapukani]} \]
    \begin{align*}
    & \text{I-PAST-for some time eat-repeatedly} \\
    & \text{‘I kept on eating’}
    \end{align*}

**Nivkh (Gilyak), Russia - Sakhalin Island** (Jakobson 1971: 80)

(12) \[\text{pöi-mu-meń-vo-ňivx} \]
    \begin{align*}
    & \text{fly-boat-helm-take person} \\
    & \text{‘aviator’}
    \end{align*}

**Phonology (sounds)**

Some languages are consonant oriented (BDFGKLMNPR etc.), other languages are vowel oriented (AEIOUÆØÅ etc.)

**Danish:** a dialect of Central Jutland (Griebel 1998:5)

(13) \[\text{Jeg er ude på øen i åen} \quad \text{[Danish]} \]
    \begin{align*}
    & \text{Jeg er ude på a ø i a å} \quad \text{[dialect]} \\
    & a æ u æ æ ø i æ å \quad \text{[phonetic transcription]} \\
    & \text{‘I am out on the island in the stream’}
    \end{align*}

**Semantics (meaning)**

**Tuyuca (adapted from Barnes as cited in Palmer 1986)**

(14) \[\text{diiga apé-wi} \]
    \begin{align*}
    & \text{soccer play-VISUAL} \\
    & \text{‘He played soccer’ (I saw him)}
    \end{align*}

(15) \[\text{diiga apé-ti} \]
    \begin{align*}
    & \text{soccer play-NON_VISUAL} \\
    & \text{‘He played soccer’ (I heard him playing)}
    \end{align*}

(16) \[\text{diiga apé-yi} \]
    \begin{align*}
    & \text{soccer play-APPARENT} \\
    & \text{‘He played soccer’ (I have evidence though I didn’t actually witness the game in any way)}
    \end{align*}

(17) \[\text{diiga apé-yigÈ} \]
    \begin{align*}
    & \text{soccer play-2\textsuperscript{nd}_HAND} \\
    & \text{‘He played soccer’ (Someone told me)}
    \end{align*}

(18) \[\text{diiga apé-hiyi} \]
    \begin{align*}
    & \text{soccer play-ASSUMED} \\
    & \text{‘He played soccer’ (It seems reasonable)}
    \end{align*}
Pragmatics (rules of language use, e.g. knowing how and when to speak to a certain person in a certain situation)

For example: indirect communication / bystander deixis

In the Caribbean, addressing remarks to (or about) a person within her/her hearing but through a “sham receiver” is a common practice. The speaker avoids responsibility for the audience's assignment of intentionality by obviously and strategically providing conflicting symbols and signs. In Africa and the Caribbean, indirect communication occurs in contexts and within norms which all members of the society recognize as appropriate or inappropriate. (Reisman 1974; Fisher 1976; Morgan 1991: 424).

Members of the Cuiva tribe of Columbia (Kerr 1977: 161): “When a conversation is mixed, women speak through their husbands to another man, and men speak through their wives to another woman.”

In Australian aboriginal cultures there are certain kin relations that require special respectful linguistic and non-linguistic behavior. These relations typically involve a man's in-laws (with the notable exception of a man's wife's sisters), especially a man's mother-in-law, with whom he has to avoid all contact (at least traditionally). For instance, in Guugu Yimidhirr society a man and his mother-in-law do not sit in one another's presence, do not look at each other, approach one another, or stand face to face. Verbal interaction with other taboo-relations such as brother-in-laws is allowed, but then the man has to use the special avoidance style, which involves the use of an entirely distinct vocabulary. Furthermore the respectful language of the Guugu Yimidhirr is characterized by “a deliberately subdued voice, drawing out words and dropping into a near whisper. At the same time it is impolite to attempt physical proximity with one's in-laws; instead one dili yirrgaalga or wurrin yirrgaalga — that is, speaks “sideways” or “crosswise,” neither facing one's interlocutor nor, if it can be avoided, addressing him or her directly” (Haviland 1979: 217, 234).
2. UNITY IN DIVERSITY: what languages have in common
(at some abstract level)

2.1. Limits on variation 1: complexity
Hawkins (1983: 83): Universal IX': “If a language has noun before genitive, then it has noun before relative clause; i.e. NG ⊃ NRel (equivelently: RelN ⊃ GN )”

If N Genitive then N Relative clause.

‘bog Vibeke’s’ ‘(den) hygning, der hæver sig i baggrunden)’

<table>
<thead>
<tr>
<th>N - Relative clause</th>
<th>Relative clause - N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun – Genitive</td>
<td>e.g. Sumerian</td>
</tr>
<tr>
<td>Genitive – Noun</td>
<td>e.g. Danish</td>
</tr>
</tbody>
</table>

QUESTION FOR LINGUISTICS: Why are there no languages that have NG and Rel N?

2.2. Limits on variation 2: relevance (semantic scope)
Word order patterns in the Noun Phrase

<table>
<thead>
<tr>
<th>Demonstrative</th>
<th>Numerical</th>
<th>Adjective</th>
<th>Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>dem</td>
<td>num</td>
<td>A</td>
<td>N</td>
</tr>
</tbody>
</table>

(19)
English: those three black dogs
Danish: de tre sorte hunde
Dutch: de drie zwarte honden

Logically there are 4x3x2x1=24 possibilities to say 'those dem three num black A dogs N':

(20)

dem num A N    dem num N A    dem N num A    N dem num A
num dem A N    num dem N A    num N dem A    N num dem A
A dem num N    A dem N num    A N dem num    N A dem num
A dem A num N  A dem N num    A N dem num    N A dem num
num A num dem  num A N dem    num N A dem    N num A dem
A num A num dem num A N dem    A N num dem    N A num dem

But only about 8 (= one third) of these possibilities occur in the languages of the world:

<table>
<thead>
<tr>
<th>Word Order</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>dem num A N</td>
<td>Alamblak, Dutch, Georgian, Hungarian, Kayardild, Ket, Nama Hottentot, Imbabura Quechua, Pipil, Tamil, Turkish</td>
</tr>
<tr>
<td>dem num N A</td>
<td>Burushaski, Guaraní, French (and other Romance languages)</td>
</tr>
<tr>
<td>dem A N num</td>
<td>Zande</td>
</tr>
<tr>
<td>dem N A num</td>
<td>Bambara</td>
</tr>
<tr>
<td>num A N dem</td>
<td>Berbice Dutch Creole, Bislama, Sranan</td>
</tr>
<tr>
<td>num N A dem</td>
<td>Basque, Hmong Njua</td>
</tr>
<tr>
<td>A N num dem</td>
<td>Sango</td>
</tr>
<tr>
<td>N A num dem</td>
<td>Oromo, Fa d'Ambu, Nubi</td>
</tr>
</tbody>
</table>
General pattern (= SYMMETRY!):

(22)  \[ \text{dem num } \text{A N A num dem} \]

\begin{align*}
\text{English:} & \quad \text{three black [dogs]} \\
& \quad \text{those [three black dogs]} \\
\end{align*}

Word order possibilities are OFTEN constrained by relevance (semantic scope): in ‘those two black dogs’ ‘black’ only relates to the noun ‘dogs’; three specifies the number of ‘black dogs’; ‘those’ specifies the location of the ‘three black dogs’.

2.3. Modeling linguistic constructions

One model (to represent the underlying structure) for all languages

More on symmetry: parallels between the underlying structure of the sentence and the noun phrase

A model of the underlying semantic structure of noun phrases (e.g. ‘those three big dogs’) and sentences (e.g. ‘He often works in the garden’).

A model is NOT a copy of something. In a model you focus in certain aspects. For example, in a model of university you can focus on the physical aspects (buildings) or the organization (network), etc.

A semantic model of a linguistic structure (noun phrase, sentence) focuses on meanings and meaning relations, not on syntax (word order), morphology (word form), phonology (sounds).

Grammatical expression of Kind, Quality, Quantity and Location in the clause

Lexical expression of Kind, Quality, Quantity and Location in the clause

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{symmetry_diagram.png}
\caption{Symmetry in the underlying structure of the clause and the NP (Rijkhoff 2005).}
\end{figure}
DESCRIPTIVE MODIFIERS (with some examples of satellites in the clause and NP).

NOTE: there is one-to-one relationship between form and function, For example, a relative clause may ‘function’ as a qualifying, quantifying, or localizing modifier in the noun phrase.

### INTERPERSONAL LEVEL (‘LANGUAGE AS EXCHANGE’)  
At this level, operators (ω, π) and satellites (τ, σ) are concerned with the interpersonal status of four kinds of entities in the World of Discourse: [a] clauses (or rather the messages contained in the clauses), [b] propositions, [c] events and [d] things.

#### CLAUSE

<table>
<thead>
<tr>
<th>Operator</th>
<th>Satellite</th>
</tr>
</thead>
<tbody>
<tr>
<td>4th order entity: Message (E)</td>
<td>π₆, σ₆</td>
</tr>
</tbody>
</table>

#### ILLOCUTION LAYER

Illocutionary modifiers: S informs A about the illocutionary status of message Eᵢ.

#### PROPOSITION LAYER

Proposition modifiers: A is informed about S’s personal assessment of / attitude towards proposition Xᵢ as regards the probability, possibility or desirability of the actual occurrence of event eᵢ.

#### NOUN PHRASE

<table>
<thead>
<tr>
<th>Operator</th>
<th>Satellite</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st order: Thing (x)</td>
<td>π₄, σ₄</td>
</tr>
<tr>
<td>2nd order: Event (e)</td>
<td>π₄, σ₄</td>
</tr>
</tbody>
</table>

Discourse modifiers: S informs A about the existential status of thing xᵢ or event eᵢ in the World of Discourse.

### REPRESENTATIONAL LEVEL (‘LANGUAGE AS CARRIER OF CONTENT’)

Descriptive modifiers specify properties of an entity in the World of Discourse in terms of Kind, Quality, Quantity, and Location.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Operator</th>
<th>Satellite</th>
<th>Operator</th>
<th>Satellite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>ω₃</td>
<td>τ₃</td>
<td>π₃</td>
<td>σ₃</td>
</tr>
<tr>
<td>Quantity</td>
<td>ω₂</td>
<td>τ₂</td>
<td>π₂</td>
<td>σ₂</td>
</tr>
<tr>
<td>Quality</td>
<td>-</td>
<td>τ₁</td>
<td>-</td>
<td>σ₁</td>
</tr>
<tr>
<td>Kind</td>
<td>ω₀</td>
<td>τ₀</td>
<td>π₀</td>
<td>σ₀</td>
</tr>
</tbody>
</table>

Figure 2. NP and clause layers at the Interpersonal and the Representational Level (Rijkhoff 2005).
3. Underspecification
3.1. Vagueness across major word classes

<table>
<thead>
<tr>
<th>Flexible Parts-of-Speech Systems</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
<th>Type 5</th>
<th>Type 6</th>
<th>Type 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>contentive</td>
<td>verb</td>
<td>non-verb</td>
<td>verb</td>
<td>noun</td>
<td>adjective</td>
<td>modifier</td>
</tr>
<tr>
<td>Rigid Parts-of-Speech Systems</td>
<td>verb</td>
<td>noun</td>
<td>adjective</td>
<td>verb</td>
<td>noun</td>
<td>noun</td>
<td>verb</td>
</tr>
</tbody>
</table>

Figure 4. Parts–of–speech system (adverb = manner adverb; based on Hengeveld 1992).

**TYPE 1:**

**Samoan** (Austronesian; Mosel & Hovdhaugen 1992: 80, 73, 74)

(23) ‘Ua mālosi le lā
PERF strong ART sun
‘The sun is strong.’ (lit. “The sun strongs.”)

(24) ‘Ua lā le aso.
PERF sun ART day
‘The sun is shining today.’ (lit. “The day suns”)

**Straits Salish** (Jelinek and Demers 1994: 718)

(25) čey=Ø  cə  swəy’qə’
work=3ABS DET man
‘He works, the (one who is a) man.’

(26) swəy’qə’=Ø  cə  čey
man=3ABS DET work
‘He is a man, the (one who) works.’

**Tagalog** (Sasse 1993b: 655)

(27) nagtatrabaho ang lalaki
work REF man
‘The man is working’

(28) lalakiang nagtatrabaho
man REF work
‘The one who is working is a man’
Type 6:
Galela (van Baarda 1908: 35)

(29) awi dohu i lalamo
his foot it be_big:PRT
‘his big foot’

Hausa (Schachter 1985: 15)

(30) mutum mai alheri / arzaki / hankali
person with kindness / prosperity / intelligence
‘a kind/prosperous/intelligent person’

Type 7:
Cayuga (Sasse 1993b: 657)

(31) a-hó-hto:’
ho-tkwe’-a’
ne:kyé h-okweh
PAST-it:to_him-become_lost it:him-wallet-be this he:it-man
‘This man lost his wallet’

The literal meaning, however, would be something like (Sasse: ibid.): “it became lost to him, it is his wallet, he is this man" or rather: "it losted him, it wallets him, the one who mans”.

Tuscarora (Mithun Williams 1976: 32)

(32) rakwá:tihs wahratkáhtho? katéskrahs
ra-kwathihs wa-hr-at-kahtho?’ ka-tesk-r-ah
M-young AOR-M-look_at-PNCT NonH-stink-SER
he_is_young he_looked_at_it it_stinks
‘The boy looked at the goat’

‘Flexible languages’: There are always disambiguating elements in the linguistic or non-linguistic context. Morphosyntactic strategies: [a] fixed word order, [b] morphological marker (Hengeveld & Rijkhoff & Siewierska 2004, Hengeveld & Rijkhoff 2005).

[a] At the sentence level all flexible languages in a 50-language sample turn out to be either (main) predicate final or (main) predicate initial (or in traditional terms that would be inappropriate for truly flexible languages: SOV or VSO). Apparently, flexible languages exploit the highly recognizable sentence-initial and sentence-final position for identificational purposes. Obviously any other position for the main predicate (such as predicate medial or ‘SVO’) would be problematic for a language without a clear noun–verb distinction if no other clues are available (see below). Significantly, there are no such constraints in languages with rigid PoS systems, which from a cross-linguistic perspective can have the main predicate in any position in the sentence.

[b] If a flexible language allows for word order variation, there is always a morphological element to mark the role of a lexeme. For example, in Samoan, placement of any referential phrase in sentence intitial position, i.e. before the predicate phrase, is accompanied by the addition of a morphological marker, the presentative particle ‘o.

Samoan (Mosel & Hovdhaugen 1992: 52, 56)

(33) ‘Ua o tamaiti i Apia
PERF go children LD Apia
‘The children have gone to Apia’
3.2. Underspecified nouns

<table>
<thead>
<tr>
<th>SPACE</th>
<th>-HOMOGENEITY</th>
<th>+HOMOGENEITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>-SHAPE</td>
<td>general noun</td>
<td>mass noun</td>
</tr>
<tr>
<td>sort noun</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+SHAPE</td>
<td>set noun</td>
<td>collective noun</td>
</tr>
</tbody>
</table>

Figure 5. Classification of nouns: six Seinsarten (Rijkhoff 2002)
[cf. AKTIONSARTEN in verb semantics]

Across languages FOUR KINDS OF NOUN (encoding FOUR DIFFERENT Seinsarten) are used to talk about concrete objects:

Dutch: singular object noun or 'count noun' [+Shape, -Homogeneity]
(33) twee boek-en [numeral+noun+plural]
two book-PL 'two books'

Oromo (Stroomer 1987:59): set noun [+Shape]
(34) gaala lamaani [noun+numeral; no plural]
camel(s)two 'two camels'

Thai (Hundius and Kölver 1983:172): sort noun [–Shape, –Homogeneity]
(35) rôm sàam khan [noun+numeral+sortal classifier]
umbrella(s)three CLF:long, handled object 'three umbrellas'

Yucatec Maya (Lucy 1992:74): general noun [–Shape]
a/one-CLF banana [numeral+general_classifier+noun]
(36) a. ‘un-tz’íit há’as’ one/a 1-dimensional banana (i.e. the fruit)'
b. ‘un-wáal há’as’ one/a 2-dimensional banana (i.e. the leaf)'
c. ‘un-kúul há’as’ one/a planted banana (i.e. the plant/tree)'
d. ‘un-kúuch há’as’ one/a load banana (i.e. the bunch)'
e. ‘um-p’íit há’as’ one bit banana (i.e. a bit of the fruit)'

SINGULAR OBJECT NOUNS, which are typically found in the European languages, are SPECIALIZED in that they can only be used to refer to a single concrete object or thing (and to more than one if provided with a plural marker). Speakers of most languages, however, employ TRANSNUMERAL NOUN types to refer to a concrete object: the same (unmarked) form can be used to refer to one or more objects in e.g. Oromo, Thai and Yucatec Maya (see above). This is so because transnumeral noun types are not deemed to have the same values for the features SPACE and HOMOGENEITY. Whereas a SINGULAR OBJECT NOUN denotes a non-homogeneous property with a spatial outline (+Shape, -Homogeneity), which matches the properties of a concrete object in the real world, the other three noun types have different values for these features (or no value for Homogeneity in the case of GENERAL NOUNS) and can thus be said to be underspecified with respect the entity they refer to.
In fact we can distinguish two kinds of underspecification or vagueness:

1. **GENERAL NOUNS** (e.g. Yucatec Maya) are vague in the sense of Cruse (1986: 51): they have a general meaning which covers two or more of the specific possibilities (singular object, mass, collective).

2. **SORT NOUNS** (e.g. Thai, Vietnamese and many other SE Asian languages) on the other hand, are used to refer to concrete objects, but they are underspecified in that their meaning definitions are believed not to include the notion of spatial boundedness or discreteness. That is to say, SORT NOUNS “purely denote concepts” (Hundius & Kölver 1983: 166; cf. also Lyons 1977: 462), which makes them very flexible for referential purposes: different classifiers coerce different meanings from the noun (Huang & Ahrens 2002: 355).

### 4.3. Underspecified speech acts: Indirect speech acts etc.

In normal human communication it is important to encode the message in such a way so as not to affront the addressee’s face.

**FACE** = the public self image that every member of society wants to claim for himself (Goffman, Brown & Levinson 1978: 66)

**POSITIVE FACE** = an individual’s desire to seem worthy and deserving of approval

**NEGATIVE FACE** = an individual’s desire to be autonomous, unimpeded by others.

A kind of mutual self-interest requires that conversational participants maintain both their own face and their interactor’s face. In this view, many verbal interactions are potential threats to face. In the continual interactive balancing of one’s own and others’ face, politeness serves to diminish potential threats.

Various linguistic practices (forms of politeness) are employed to minimize threats to

- **POSITIVE FACE** (one’s positive self-esteem): strategies to make the addressee feel good about himself;

- **NEGATIVE FACE** (one’s freedom to act; freedom from imposition or constraint): strategies (used by the speaker) to gloss over/minimize imposition.

### Degrees of indirectness: Imperatives

Degrees of manipulative power are reflected in the range of manipulative constructions (from canonical imperative to deferent request; Givón 1990: 807-8):

1. Get up!
2. Get up, will you.
3. Would you please get up?
4. Would you mind getting up?
5. Do you think you could get up?
6. Would you mind if I asked you to get up?
7. Wouldn’t it be nice if you could perhaps get up?

### Devices for weakening manipulative strength

- increased length
- using question form
- overt mentioning of the manipulee pronoun
- use of irrealis modality on the verb
- use of negation
- embedding under modality or cognition verbs.