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) *Toìkyoì eki   kara denìsha de ichijikanì kurai nishi e*
    Tokyo   station   from tram   by one hour   about west toward
    14   13   12   11 10   9   8

    itta tokoro ni Kamakura to   iu   machi ga arimasu
    go   place   at Kamakura   "Quote" called town   OBJ exists
    7   6   5   4   3   2   1

(2) *There   is a town called Kamakura at a place (you can reach)*
             1   2   3   4   5   6   7

    going west about one hour by tram from Tokyo Station.
    8   9   10   11 12   13   14

☞ **Syntax** (sentence structure): <u>**S**</u>**ubject** - <u>**V**</u>**erb** - <u>**O**</u>**bject** 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)
(3) SOV <	SVO <	VSO <	VOS	< OVS	< OSV	

<u>Japanese</u>:	**Subject**	**Object**	**Verb**
(4)	*Taroo ga*	*tegami o*	*kaita$_V$*
	Taroo SUBJ	letter OBJ	wrote
	'Taroo wrote a letter'		

<u>Danish</u>:	**Subject**	**Verb**	**Object**
(5)	*Vibeke*	*skrev*	*et brev*

<u>Welsh</u>:	**Verb**	**Subject**	**Object**
(6)	*Lladdodd*	*y ddraig*	*y dyn*
	killed	the dragon	the man'
	'The dragon killed the man'		

<u>Malagasy</u>:	**Verb**	**Object**	**Subject**
(7)	*Nahita*	*ny vehivavy*	*ny mpianatra*
	saw	the woman	the student
	'The student saw the woman'		

<u>Hixkaryana</u>:	**Object**	**Verb**	**Subject**
(8)	*Toto*	*yonove*	*kamara*
	man	he-ate-him	jaguar
	'The jaguar ate the man'		

☞ **Morphology (word structure)**: separate words (isolating) or all-in-one (polysynthetic)

Bejing Chinese, China (Chrystal 1987: 293)

(9)   *Wǒ   mǎi   júzi   chī*
        I   buy   orange   eat
        'I bought some oranges to eat'

Swahili, Kenya (Chrystal 1987: 293)

(10)   mimi   ni-na-ku-penda   wewe   [*mini ninakupenda wewe*]
        me   I-PRESENT-you-love   you
        'I love you'

Tiwi, Australia (Chrystal 1987: 293)

(11)   ngi-rru-unthing-apu-kani   [*ngirruunthingapukani*]
        I-PAST-for_some_time-eat-repeatedly
        'I kept on eating'

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

(12)   *pəi-mu-meñ-vo-ñivx*
        fly-boat-helm-take-person
        'aviator'

☞ **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)   Jeg   er   ude   på   øen   i   åen   [Danish]
        Jeg   er   ude   på   a ø   i   a å   [dialect]
        a   æ   u   å   æ ø   i   æ å   [phonetic transription]
        'I am out on the island in the stream'

☞ **Semantics (meaning)**

Tuyuca (adapted from Barnes as cited in Palmer 1986)

(14)   *díiga   apé-wi*
        soccer   play-VISUAL
        'He played soccer' (I saw him)

(15)   *díiga   apé-ti*
        soccer   play-NON_VISUAL
        'He played soccer' (I heard him playing)

(16)   *díiga   apé-yi*
        soccer   play-APPARENT
        'He played soccer' (I have evidence though I didn't actually witness the game in any way)

(17)   *díiga   apé-yigÈ*
        soccer   play-2^nd_HAND
        'He played soccer' (Someone told me)

(18)   *díiga   apé-hiyi*
        soccer   play-ASSUMED
        'He played soccer' (It seems reasonable)

☞   **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 *diili 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 (equivilently: RelN ⊃ GN)"

If   *N*   **Genitive**   then   *N*   **Relative clause.**
        '*bog* **Vibekes'**   '(den)   *bygning*,   **der hæver sig i baggrunden**)'

	N - Relative clause	Relative clause - N
Noun – Genitive	e.g. Sumerian	***NO LANGUAGES***
Genitive – Noun	e.g. Danish	e.g. Japanese

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

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

		Demonstrative	**Num**eral	**A**djective	**N**oun
		dem	num	A	N
(19)	English:	those	three	black	dogs
	Danish:	de	tre	sorte	hunde
	Dutch:	die	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
dem A num N	dem A N num	dem N A num	N dem A num
num A dem N	num A N dem	num N A dem	N num A dem
A num dem N	A num 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:

	Word Order	Language
(21)	dem num A N	Alamblak, Dutch, Georgian, Hungarian, Kayardild, Ket, Nama Hottentot, Imbabura Quechua, Pipil, Tamil, Turkish
	dem num N A	Burushaski, Guaraní, French (and other Romance languages)
	dem A N num	Zande
	dem N A num	Bambara
	num A N dem	Berbice Dutch Creole, Bislama, Sranan
	num N A dem	Basque, Hmong Njua
	A N num dem	Sango
	N A num dem	Oromo, Fa d'Ambu, Nubi

General pattern (= SYMMETRY!):

(22)	**dem**	**num**	**A**	**N**	**A**	**num**	**dem**
English:			black	[dogs]			
		three	[black	dogs]			
	those	[three	black	dogs]			

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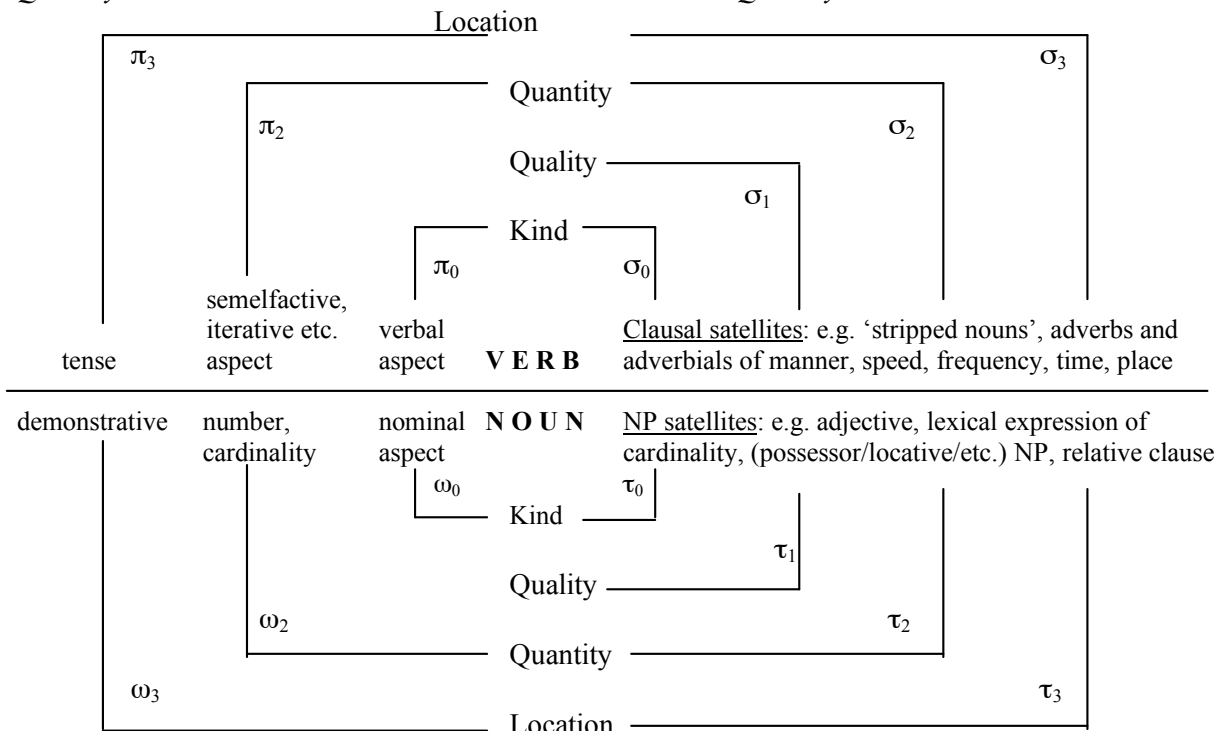
A model of the underlying semantic structure of noun phrases (e.g. 'those three big dogs') and sentences (e.g. 'He often works in the garden'').

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

Location

π_3 Quantity σ_3

π_2 σ_2

Quality

σ_1

Kind

π_0 σ_0

tense	semelfactive, iterative etc. aspect	verbal aspect	**V E R B**	Clausal satellites: e.g. 'stripped nouns', adverbs and adverbials of manner, speed, frequency, time, place
demonstrative	number, cardinality	nominal aspect	**N O U N**	NP satellites: e.g. adjective, lexical expression of cardinality, (possessor/locative/etc.) NP, relative clause

ω_0 τ_0

Kind

τ_1

Quality

ω_2 τ_2

Quantity

ω_3 τ_3

Location

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

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

Figure 2. Symmetry in the underlying structure of the clause and the NP (Rijkhoff 2005).

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

	Operator	Satellite
4th order entity: Message (E)		
ILLOCUTION LAYER	π_6	σ_6

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

	Operator	Satellite
3rd order entity: Proposition (X)		
PROPOSITION LAYER	π_5	σ_5

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

NOUN PHRASE

	Operator	Satellite		Operator	Satellite
	1rst order: Thing (x)			2nd order: Event (e)	
ENTITY LAYER	ω_4	τ_4		π_4	σ_4

Discourse modifiers: S informs A about the existential status of thing x_i or event e_i 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.

Layer	Operator	Satellite	Operator	Satellite
Location	ω_3	τ_3	π_3	σ_3
Quantity	ω_2	τ_2	π_2	σ_2
Quality	-	τ_1	-	σ_1
Kind	ω_0	τ_0	π_0	σ_0

Figure 2. NP and clause layers at the Interpersonal and the Representational Level
(Rijkhoff 2005).

3. Underspecification
3.1. Vagueness across major word classes

FLEXIBLE	Type 1	contentive			
PARTS-OF-SPEECH	Type 2	verb		**non-verb**	
SYSTEMS	Type 3	verb	noun	**modifier**	
RIGID	Type 4	verb	noun	adjective	adverb
PARTS-OF-SPEECH	Type 5	verb	noun	adjective	
SYSTEMS	Type 6	verb		noun	
	Type 7	verb			

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'qˊə'=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:

<u>Galela</u> (van Baarda 1908: 35)
(29)   *awi   dohu   **i**   **lalamo***
     his   foot   it   be_big:PRT
     'his big foot'

<u>Hausa</u> (Schachter 1985: 15)
(30)   *mutum   **mai   alheri***   /   ***arzaki***   /   ***hankali***
     person   with   kindness   /   prosperity   /   intelligence
     'a kind/prosperous/intelligent person'

Type 7:

<u>Cayuga</u> (Sasse 1993b: 657)
(31)   *a-hó-hto:'*   *ho-tkwe't-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".

<u>Tuscarora</u> (Mithun Williams 1976: 32)
(32)   *rakwá:tihs   wahratkáhtho?*   *katéskrahs*
    ra-kwatihs   wa-hr-at-kahtho-?   *ka-teskr-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'

(34) `***O*** *le* *maile sa* *fasi* *e* *le* *teine*
     PRES ART dog PAST hit   ERG ART girl
     'The dog was hit by the girl'

3.2. Underspecified nouns

SPACE	-HOMOGENEITY	+HOMOGENEITY
-SHAPE	**g e n e r a l n o u n**	
	sort noun	mass noun
+SHAPE	**s e t n o u n**	
	singular object noun	collective noun

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   săam   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 type 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.