Pāṇini's Razor

Paul Kiparsky

Paris, Oct. 29, 2007

Occam's Razor

- Entia non sunt multiplicanda praeter necessitatem.
  Entities should not be multiplied beyond necessity. (Attributed to William of Occam.)
- Frustra fit per plura quod potest fieri per pauciora.
  It is unnecessary to do with many things what can be done with fewer. (Occam's words.)
- Nature does nothing in vain, and more is in vain when less will serve; for Nature is pleased with simplicity, and affects not the pomp of superfluous causes. (Newton's version.)
- Select the theory that introduces the fewest assumptions and postulates the fewest entities. (A modern restatement.)

What does Occam's Razor really do?

When is an entity "necessary"?

- Many of Pāṇini's *saṃjñās* (technical terms) could be eliminated and replaced by their definitions.
- So are they necessary? Would Occam's Razor shave them off?
- Many of Pāṇini's *saṃjñās* are rule-generated but never used. E.g. only one fifth of the 200+ defined *pratyāhāras* are actually used. Should Pāṇini have complicated his grammar to exclude them?
- The idea is that theoretical concepts are justified if the contribute to the explanation of phenomena. It is not always clear how to apply this criterion.

Chomsky's Razor

- Internally to the theory of language, devise a notational system for grammars which converts simplicity into brevity.
- . . . "we are not interested in reduction of the length of grammars for its own sake. Our aim is rather to permit just those reductions in length which reflect real simplicity, that is, which will turn simpler grammars (in some partially understood, presystematic sense of this notion) into shorter grammars." (Chomsky 1955: 118)

Early generative grammar: two kinds of "simplicity"

A theory of language provides

- a fixed formalism for grammars, e.g. phonological rules in distinctive feature notation,
- an evaluation measure (the "simplicity criterion") which selects for a given language the optimal grammar that is consistent with this formalism.
- Influenced by Nelson Goodman'n writings on philosophy of science.

Theories of language are judged by general (unformalized) scientific criteria, such as conceptual elegance, simplicity, and predictive success.

The simplicity criterion

- Applies globally to the whole grammar.
- The lexicon is part of the grammar.
- Rules are written in feature notation
- The grammar is minimally specified (minimize redundancy).
- Intuitive idea: the simplicity criterion requires rules to "pay their way": they must "save" more feature specifications than they "cost".

The simplicity criterion is an *empirical hypothesis*

- Part of a *theory of language*,
- which makes predictions about grammars,
- and about language acquisition.
- So data about languages and their acqusition can in principle support or falsify it.

The simplicity criterion at work

- Data: *brick* (actual word), **blick* (possible word — an accidental gap), **bnick* (impossible word, systematic gap)
- A formal explanation of the data (Halle 1962)
  - A simple rule C → [–nasal] / #C_ excludes */bnik/, allows many features to be left unspecified in the lexicon.
  - Complex rule C → [–lateral] / b_ik needed to exclude */blik/, saves only one feature specification the lexicon.
  - By the simplicity criterion, only the first rule is correct.
  - So a theory which expresses phonological rules in distinctive feature notation *and* which incorporates the simplicity criterion distinguishes correctly between accidental and systematic gaps.

Pāṇini's Razor: *lāghava*

- *ardhamātrālāghavena putrotsavaṃ manyante vaiyākaraṇāḥ*

Pāṇini's Razor

- The simplicity metric applies to the *whole system*: rules *(Aṣṭādhyāyī)* and data structures for phonology and lexicon *(Śivasūtras, Gaṇapāṭha, Dhātupāṭha)*.
- The metalanguage (technical terms and some conventions of rule application) are defined *within* the grammar itself.
- Shortest grammar = shortest theory.

Consequences of Pāṇini's Razor

- Abstract categories such as *kārakas, lakāras,. . .*,
- principles governing rule interaction: blocking *(utsarga-apavāda)*, *anuvṛtti, adhikāra*, the *siddha*-principle,
- abbreviations for recurrent arbitrary classes: *ghi, ghu, bha, ārdhadhātuka. . .*

The inductive value of Pāṇini's Razor

- The core categories of Pāṇini's Sanskrit grammar recur in widely divergent languages.
- Also principles like Blocking *(utsarga/apavāda, sāmānya/viśeṣa)* and the *siddha*-principle.
- Arbitrary classes are grammatical realities (declensions, conjugations, etc.).
- Conclusion: systematic application of Pāṇini's Razor to one language brings out concepts and principles that apply to other languages. This suggests that Pāṇini's Razor is a sound methodological basis for linguistics.

Pāṇini's Razor

- *Not* intended as an empirical "theory of language", or as an epistemological or philosophical principle.
- But we may anachronistically *construe* it as such,
- in which case it is interestingly different from generative grammar, possibly superior.
- Cf. modern ideas on induction and algorithmic complexity (Minimum Description Length, Kolmogorov Complexity).
- Hypothesis: "A universal measure of algorithmic complexity is enough to provide an explanation for properties of grammars." (Goldsmith 2007)

Kolmogorov Complexity

- The Kolmogorov Complexity of a theory is the length of its minimal description.
- More precisely, it is the size in bits of the shortest binary program to compute a description of the theory on a universal computer. (Li & Vitányi 1997: 319)
- Among all hypotheses consistent with the data the one with the least Kolmogorov complexity is the most likely one. (Li & Vitányi 1997: 319)
- Two strands: (1) induction, statistical prediction, (2) data compression, Formal Concept Analysis.

1.	a	i	u			Ṇ
2.				ṛ	ḷ	K
3.		e	o			Ñ
4.		ai	au			C
5.	h	y	v	r		Ṭ
6.					l	Ṇ
7.	ñ	m	n̄	ṇ	n	M
8.	jh	bh				Ñ
9.			gh	ḍh	dh	Ṣ
10.	j	b	g	ḍ	d	Ś
11.	kh	ph	ch	ṭh	th	
			c	ṭ	t	V
12.	k	p				Y
13.		ś	ṣ	s		R
14.	h					L

The Śivasūtras

The simplicity principle, by selecting the shortest grammar, determines both the ordering of sounds and the placement of markers among them.

- To be grouped together in a *pratyāhāra*, sounds must make up a continuous segment of the list.
- Economy requires making the list as short as possible, which means avoiding repetitions of sounds, and using as few markers as possible.
- Consequently, if class A properly includes class B, the elements shared with B should be listed last in A; the marker that follows can then be used to form *pratyāhāras* for both A and B.

The Śivasūtras are optimal

Wiebke Petersen (2004), "A mathematical analysis of Pāṇini's Śivasūtras", *Journal of Logic, Language, and Information* 13: 471-489, and "How Formal Concept Lattices solve a problem in ancient linguistics" (2006).

- Minimal sound inventory: predictable sounds are omitted.
- The list of sounds cannot be further shortened (in particular, *h* must be listed twice). (Proved by Petersen.)
- Minimal necessary *anubandhas*: none could be removed or added without complicating the grammar.
- Optimally arranged: no rearrangement of sounds and/or anubandhas results in a simpler grammar.

Pāṇini's version of Occam's Razor

- The metalanguage of the Aṣṭādhyāyī is itself a semi-generative subsystem.
- E.g. the grammar defines many *pratyāhāras* which are never used.
- 8.2.76 **rvor upadhāyā dīrgha ikaḥ** lengthens *i*, *u* before *pada*-final *r*, *v*. The class *r*, *v* could also have been specified with the already defined (but never used) *pratyāhāra* **vaṬ*.
- It could even have been specified as **yaṬ* (since *y* doesn't occur in the relevant environment).
- Pāṇini minimizes the number of *pratyāhāras*. He uses only the ones that his simplicity principle forces him to.

Specificity: systematic avoidance of overgeneralization

- 7.4.61 *śarpūrvāḥ khayaḥ* (60 *śeṣaḥ)* 'unvoiced stops *(khaY)* after fricatives *śaR* remain' (e.g. *tiṣṭhāsati*).
- This rule could have been vacuously generalized to apply after the more inclusive set of sounds *śaL (ś, ṣ, s, h)* rather than after just *śaR*, for there are no *h*+stop clusters.
- Pāṇini avoids vacuous overgeneralization. Among equally simple formulations, systematically chooses the most specific one — if possible, one which covers only the actually occurring cases.

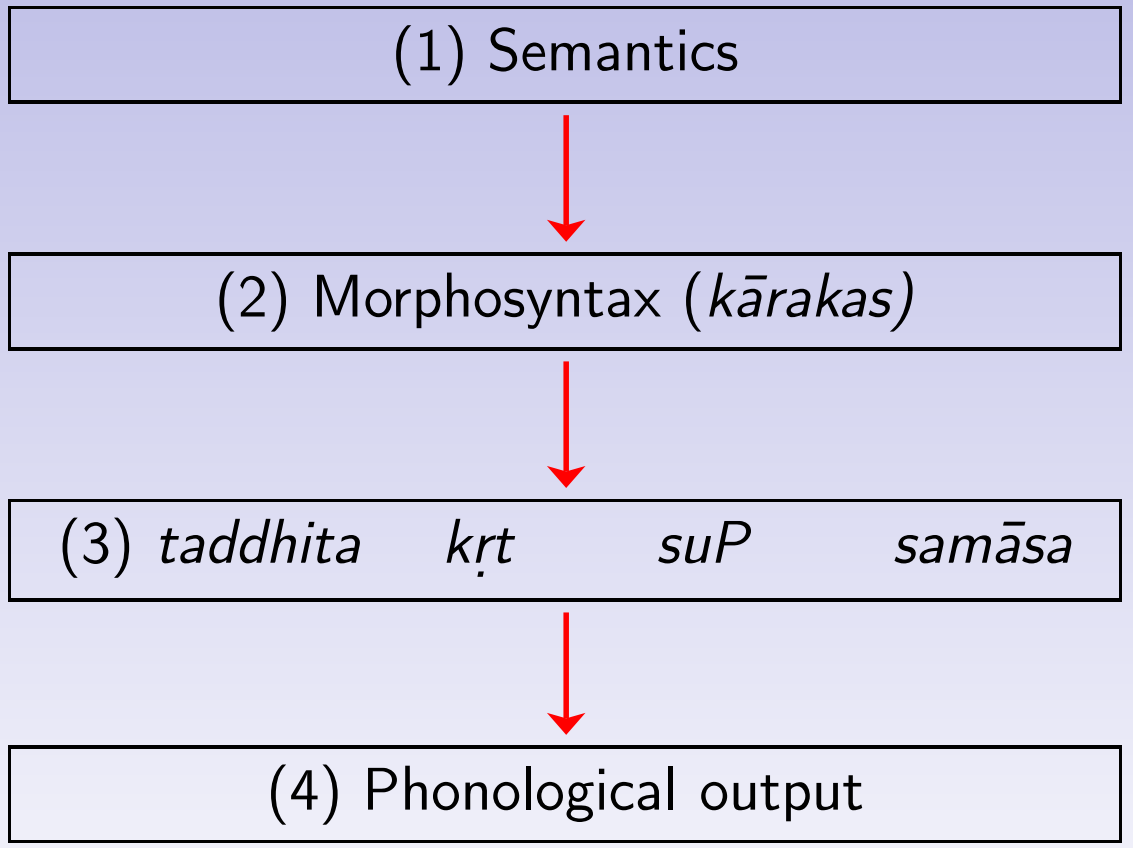
Pāṇini's hierarchy of economy principles

- Both Occam's Razor and Specificity are conservative curbs on the overgeneralization that Pāṇini's Razor otherwise enforces.
- But what about their importance relative to each other?
- 1.1.48 **eca ig ghrasvādeśe** '*e, o, ai, au* shortens to *i, u, ṛ, ḷ*'. Why include ṛ, ḷ, when this case never arises? (N.B. *iK* is not continued by *anuvṛtti*).
- Vacuous overgeneralization could have been avoided by the already defined but unused *pratyāhāra* **iṆ* = *i, u* (cf. *aṆ* = *a, i, u*). This shows that Occam's Razor outranks Specificity.

Conclusion

- Ranking of economy principles:
  Pāṇini's Razor $\gg$ Occam's Razor $\gg$ Specificity.
- From a modern perspective, this offers an interesting approach to the problem of induction (and of language acquisition in particular), where the problem is to find a learning mechanism that steers between overgeneralization and overfitting to data.

Levels of the grammar



Taddhita suffixes

Suffix_1

Suffixes which block Suffix_1 in all of its meanings:

$\text{Suffix}_{1'}$ with stem classes $X_{1'}$, $Y_{1'}$, ...

$\text{Suffix}_{1''}$ with stem classes $X_{1''}$, $Y_{1''}$, ...

etc.

Meaning_{1a} of Suffix_1, $\text{Suffix}_{1'}$, $\text{Suffix}_{1''}$, ...

Suffixes which block Suffix_1 in Meaning_{1a}:

$\text{Suffix}_{1a'}$ with stem classes $X_{1a'}$, $Y_{1a'}$, ...

$\text{Suffix}_{1a''}$ with stems $X_{1a''}$, $Y_{1a''}$, ...

etc.

Meaning_{1b} of Suffix_1, $\text{Suffix}_{1'}$, $\text{Suffix}_{1''}$, ...

Suffixes which block Suffix_1 in Meaning_{1b}:

$\text{Suffix}_{1b'}$ with stem classes $X_{1b'}$, $Y_{1b'}$, ...

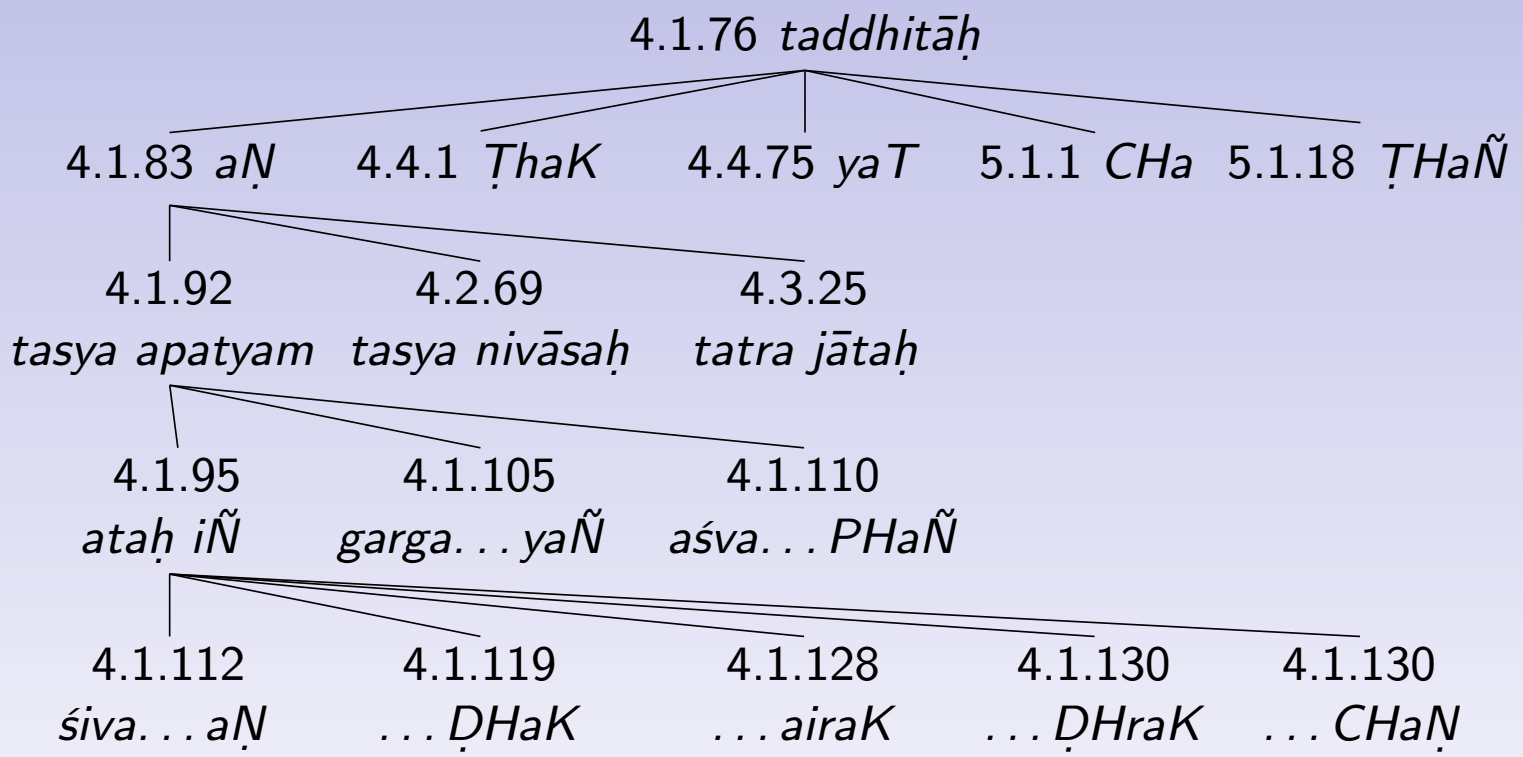
etc.

(Repeat for Suffix_2, Suffix_3, ...)

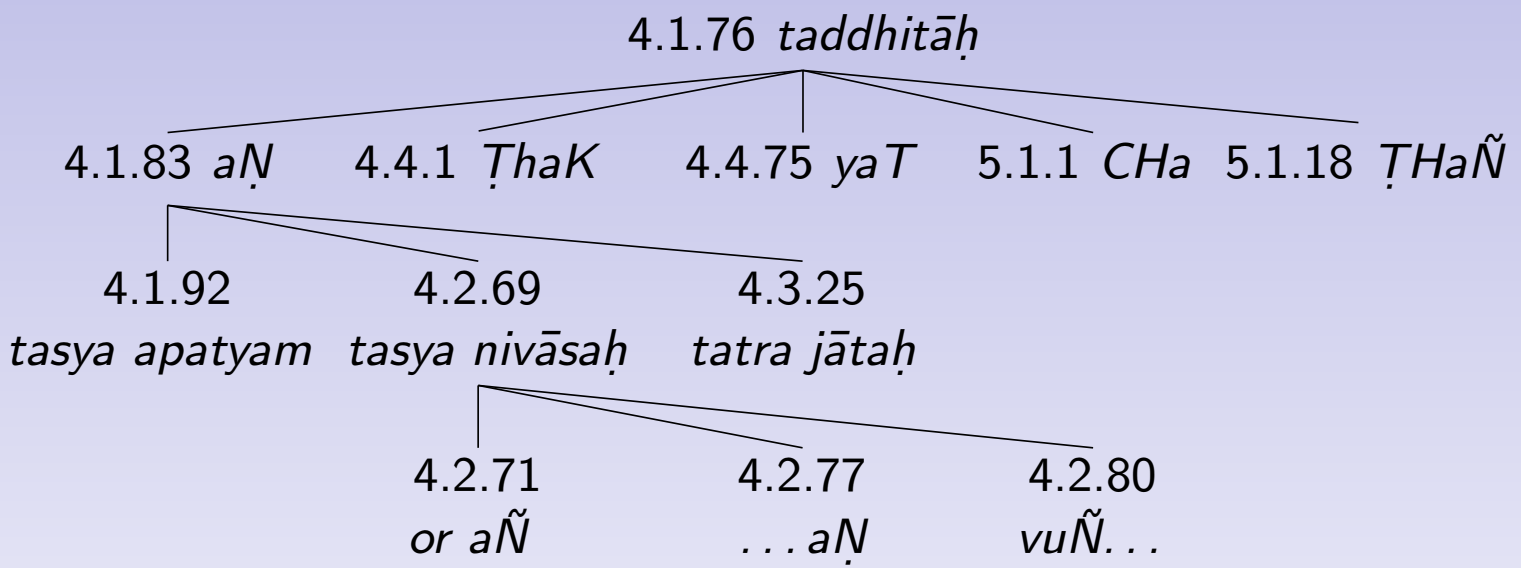
The *taddhita* section as an inheritance hierarchy

Ashwini Deo, "Derivational morphology in inheritance-based lexica: Insights from Pāṇini" (*Lingua* 2007.

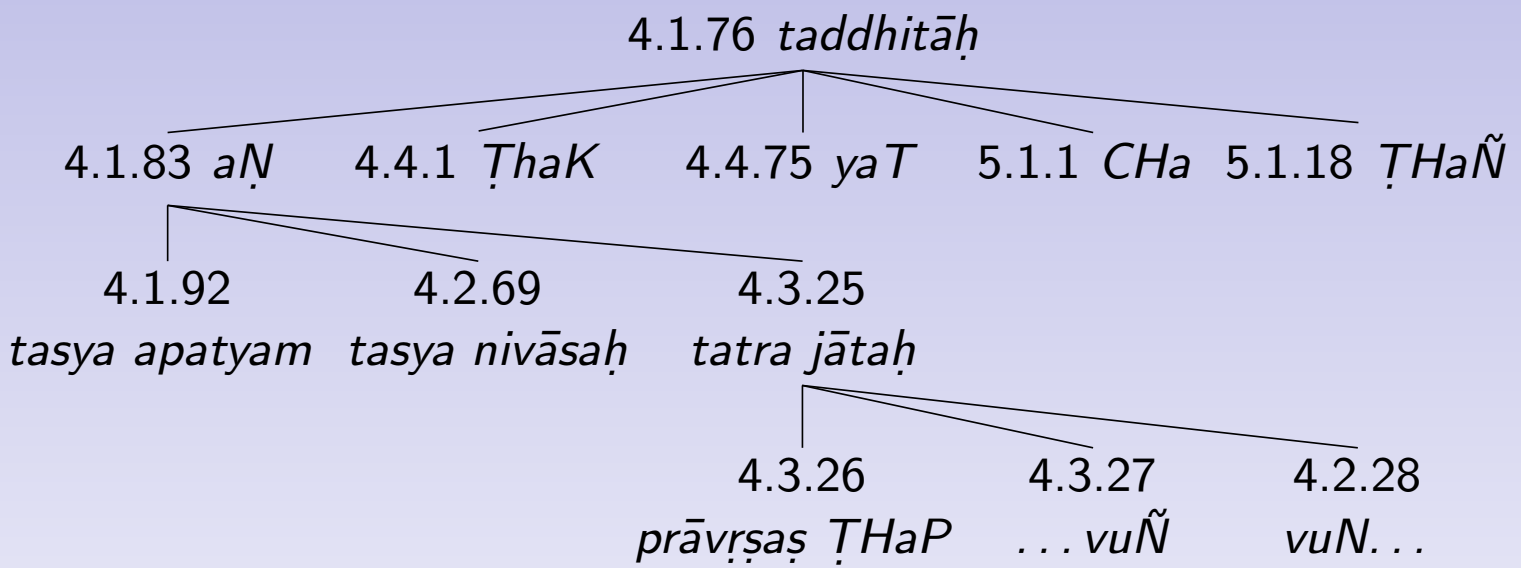
Taddhitas


4.1.76 taddhitāḥ
4.1.83 aṆ
4.4.1 ṬhaK
4.4.75 yaT
5.1.1 CHa
5.1.18 ṬHaÑ
4.1.92
tasya apatyam
4.2.69
tasya nivāsaḥ
4.3.25
tatra jātaḥ
4.1.95
ataḥ iÑ
4.1.105
garga. . . yaÑ
4.1.110
aśva. . . PHaÑ
4.1.112
śiva. . . aṆ
4.1.119
. . . ḌHaK
4.1.128
. . . airaK
4.1.130
. . . ḌHraK
4.1.130
. . . CHaṆ

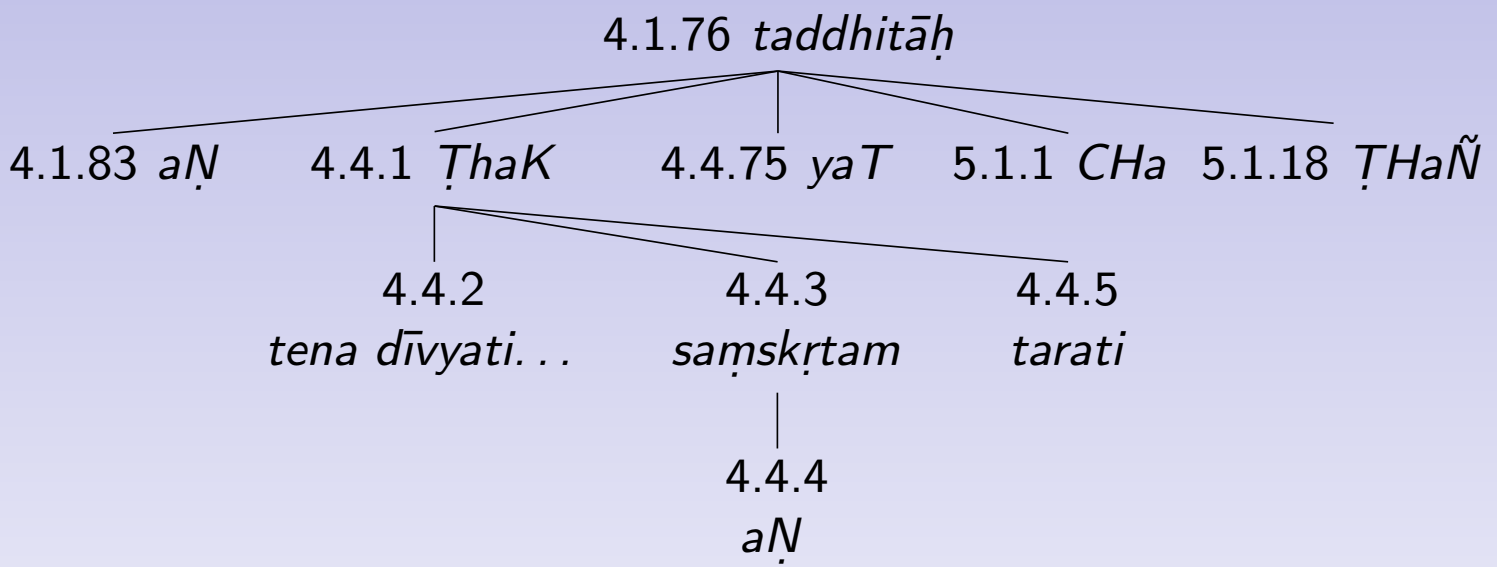
Taddhita suffixes



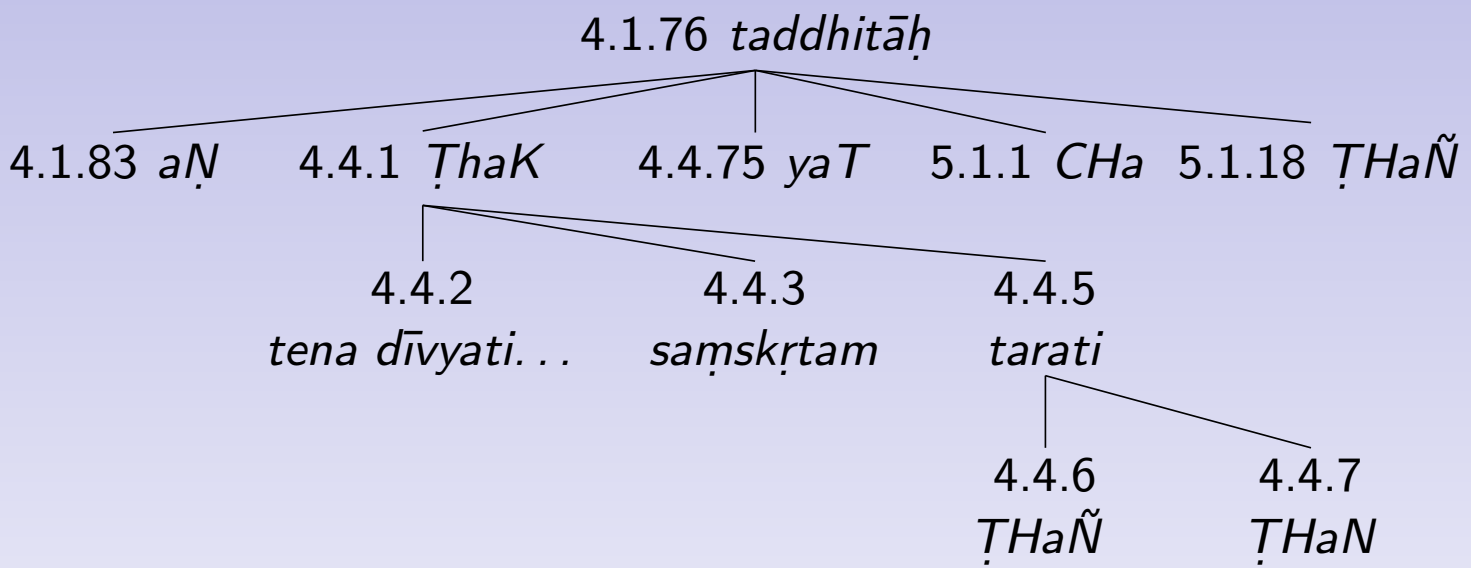
Taddhita suffixes



Taddhita suffixes



Taddhita suffixes



Types of inheritance hierarchies

Kilbury, James, Wiebke Petersen and Christof Rumpf. 2006. Inheritance-based models of the lexicon. In: Wunderlich, Dieter: *Advances in the Theory of the Lexicon*. Berlin: Mouton.

- monotonic vs. defaults
- single inheritance vs. multiple inheritance
- *Taddhitas* are a single inheritance hierarchy with defaults

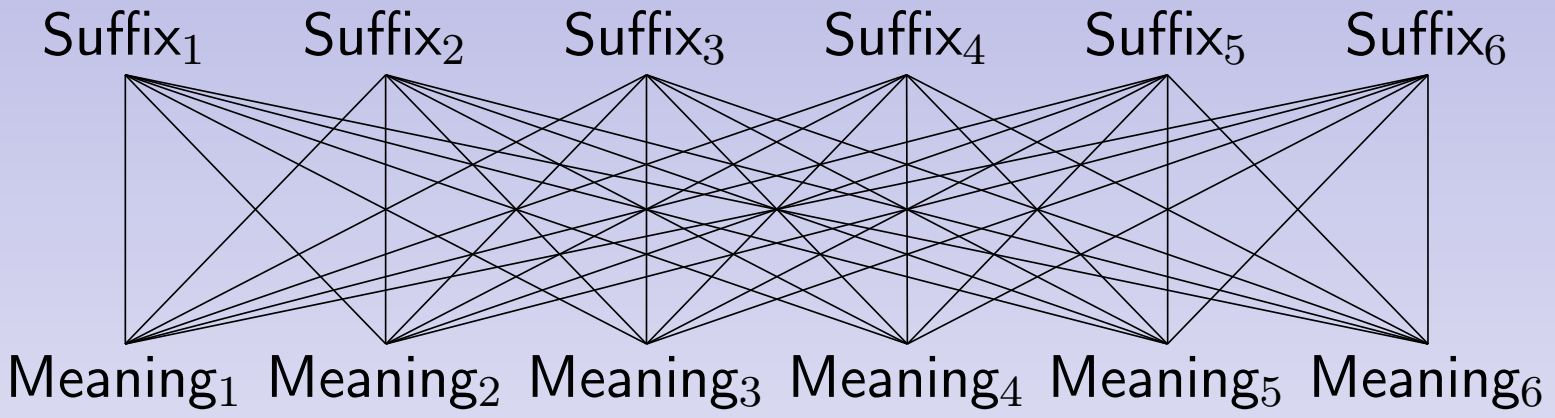
Advantages of default inheritance

- Simplifies the grammar.
- Defaults express the distinction between regular and irregular formations.
- Defaults account for productivity: the "elsewhere" option is productive.

Advantages of single inheritance

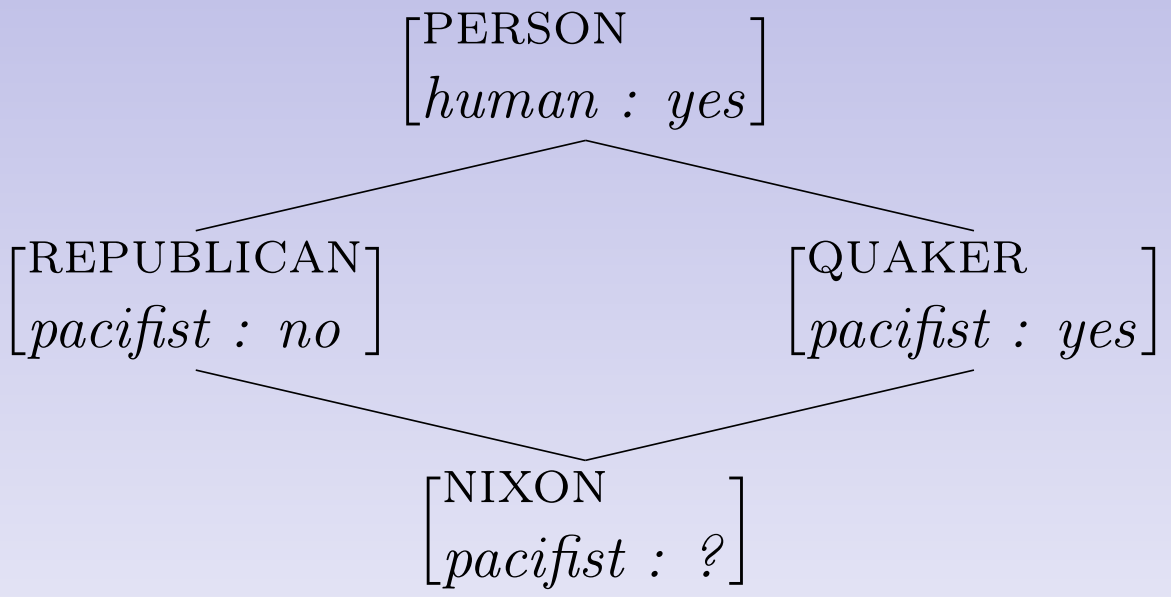
- Simpler grammars.
- More restrictive grammars.
- Formally tractable grammars: avoid problems of multiple inheritance with defaults.

Single inheritance excludes arbitrary polysemy



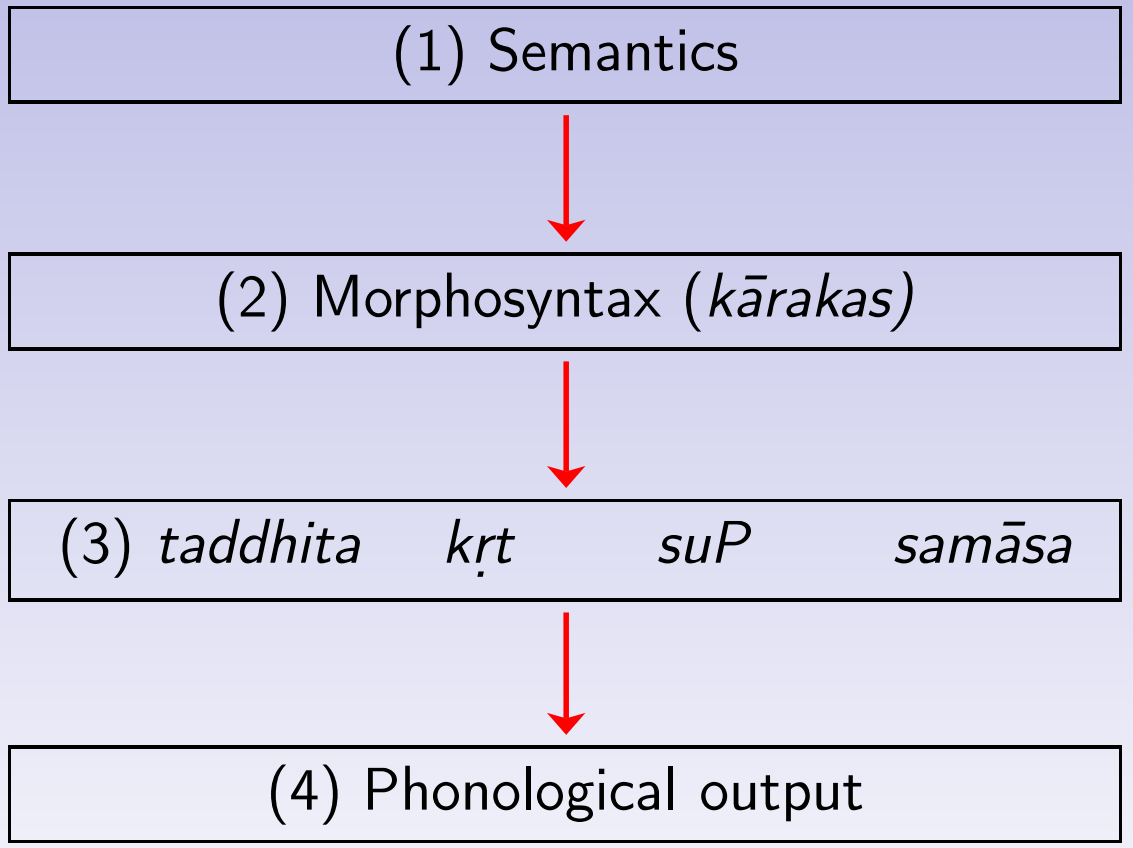
- Single inheritance hierarchies predict that every morpheme has a "basic meaning".
- Some modern morphological theories adopt multiple inheritance.

The Nixon diamond (Touretzky 1986)



- Multiple inheritance with defaults results in indeterminacy.

Levels of the grammar



Directionality

- The derivation starts from meaning
- and builds up a complete interpreted sentence.

Directionality

- The mapping between levels may depend on information from earlier levels, but not on information from later levels.
- Only morphology and phonology allow destructive (non-monotonic) operations, e.g. deletion and replacement.

Directionality

- Phonology can depend on meaning,
  6.2.48 **pratyabhivāde 'śūdre**
  'In a response to a respectful greeting, except to a Śūdra, [the last vowel is high-pitched and extra-long]
- on morphosyntax
  6.2.48 **trtīyā karmaṇi**
  áhihataḥ 'killed by a snake' (Agent)
  rathayātáḥ 'traveled by cart' (Instrument)
- on morphology

Directionality in processing

The mapping between levels may depend on information from earlier levels, but not on information from later levels.

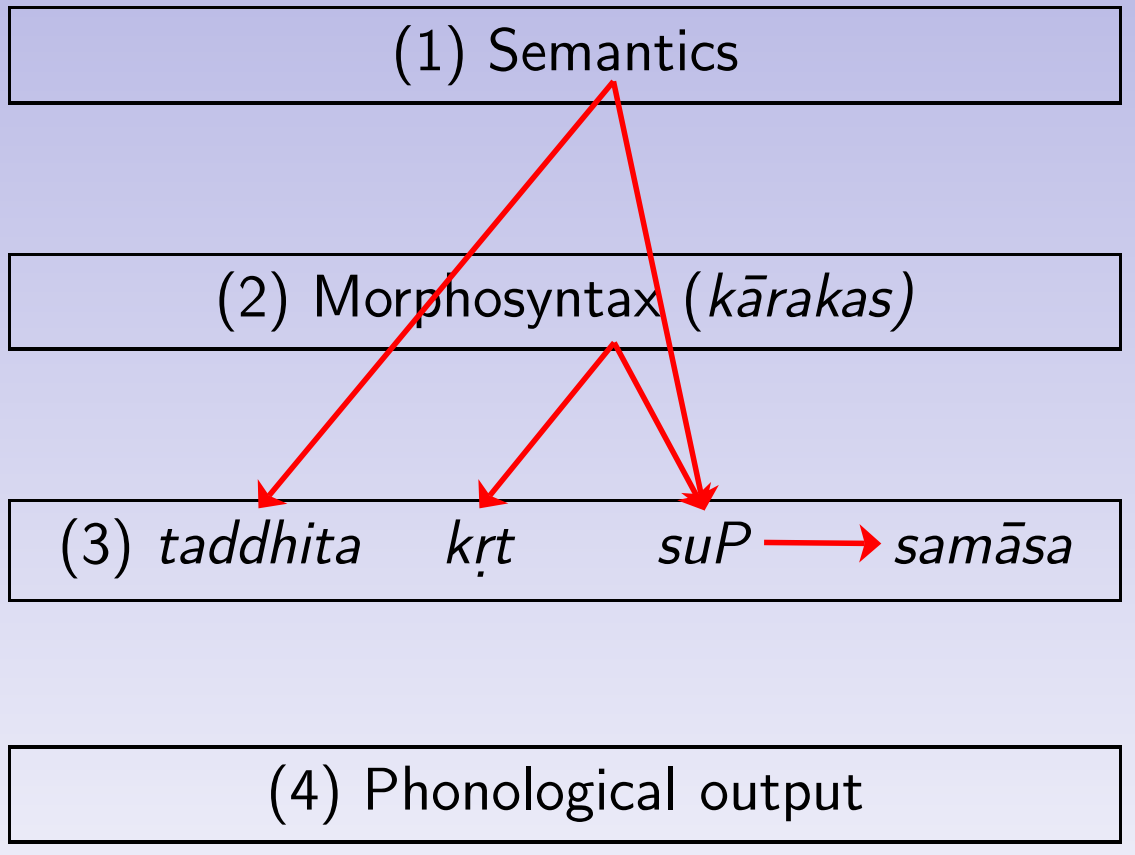
- Morphosyntax can depend on meaning
- but not on morphology or phonology.

Directionality in processing

- The higher the level, the less the computation needs to know.
- The phonology needs information about all levels.
- Therefore, top-to-bottom processing of rules is most economical.

The division of labor (Joshi 2001)

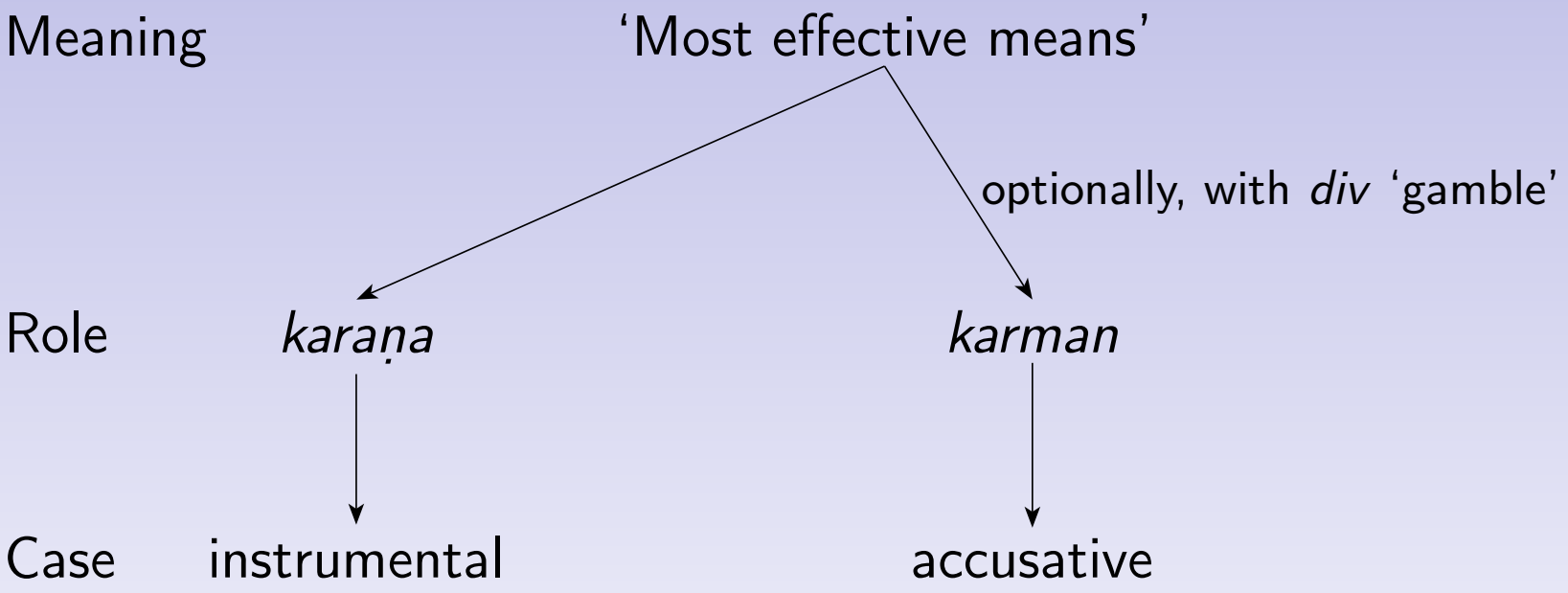
- primary derivation *(krts)* based on morphosyntax
- secondary derivation *(taddhitas)* based on semantics
- compounding *(samāsa)* based on case morphology


(1) Semantics
(2) Morphosyntax (*kārakas)*
(3) *taddhita* *kṛt* *suP* → *samāsa*
(4) Phonological output

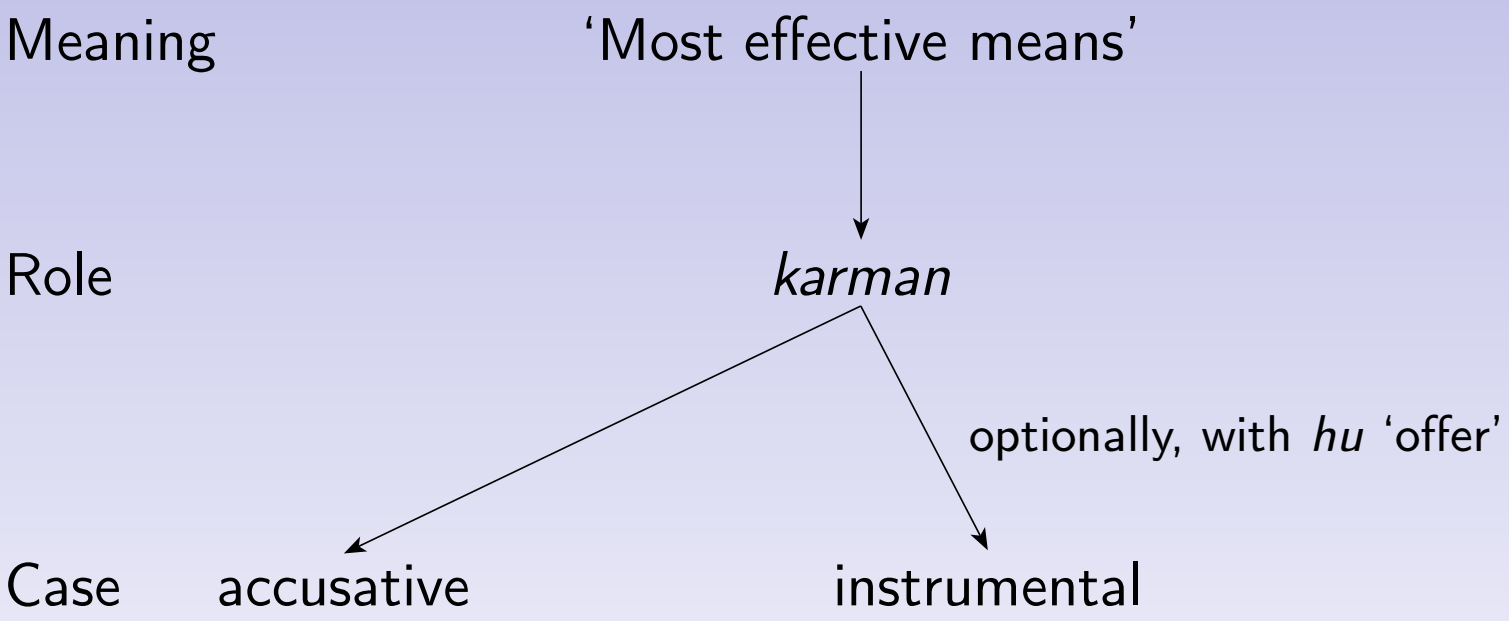
Accusative/Instrumental case variation

- *akṣān dīvyati / akṣair dīvyati* ‘he plays dice’
- *ājyam juhoti / ājyena juhoti* ‘he makes an offering of (with) ghee’

One is treated at the *kāraka* level



The other is treated at the morphological level



Why not the other way round?

Why the accusative with *div* is a *karman*

- Argument 1: passives
  akṣāḥ dīvyante (*akṣān dīvyate) ‘dice are played’
- Argument 2: genitive Goal/Patient (action nominals)
  akṣāṇāṃ devanam ‘playing (of) dice’ (*paraśunāṃ chedanam ‘cutting with axes’)

Why not the other way round?

Why the instrumental with *hu* is not a *karaṇa*

- Argument 1: *kṛt* suffixes
  havana- (1) ‘act of pouring the oblation, (2) ‘ladle’, (3) *‘oblation’ (the thing poured)
- Argument 2: prohibition of two *karaṇas*
  *dātreṇa paraśunā chinatti ‘he cuts with a sickle with an axe’
  srucā ajyena juhvati ‘he pours ghee with a spoon’

Trivial simplification

- Pointless abbreviation
  - 7.1.2 āyaneyīnīyiyaḥ phaḍhakhachaghāṃ pratyayādīnām *-phaK* → *-āyanaK*
  - Maximize sandhi in order to save syllables 3.4.72 gatyarthākarmakaśliṣaśīṅsthāsavasajanaruhajīryatibhyaḥ
- Suppose these are spurious generalizations. Does this show a flaw in Pāṇini's Razor?
- Not really. These are harmless side effects. Pāṇini's Razor works "blindly", and need not yield a generalization in every case. If data is limited, no method of inductive reasoning is guaranteed to distinguish accidental generalizations from real ones.

Globality of evaluation

Any new datum could affect the analysis of the whole system. A trivial example: the marker *C*.

- 5.2.26 **tena vittaś cuñcupcaṇapau**
  'The suffixes *-cuñcuP* and *-caṇaP* are added in the meaning "famous for X"'.

Examples

vidyācuñcu, *vidyācana* 'famous for learning'.

- *-cela*, *-cīra* (6.2.126-127), *-cara* (5.3.53), *jāhaC* (5.2.24), *jātīyaR* (5.3.69), *-ṭīṭaC* (5.2.31).
- Problem: 1.3.7 **cuṭū** designates the initial consonants of these suffixes as markers; they should be deleted by 1.3.9 **tasya lopaḥ**.
- 5.2.26 **tena vittaś ycuñcupycaṇapau**

What happened?

- Initially, the grammar had no suffixes beginning with palatals and retroflexes.
- This gap is exploited for *anubandhas* (1.3.7 **cuṭū**).
- Then marginal *taddhita* suffixes *-cuñcuP*, *-caṇaP*, etc. were discovered and retrofitted into the grammar after the system of markers had already been decided upon.
- By then it was too late to revise the grammar because the anubandhas *C* and *Ṭ* figure in hundreds of rules of the grammar and are attached to major suffixes such as *-CaN̄*, *-CiṆ*, *-CvI*, *-Ṭā*, and *-ṬāP*.
- Reworking the grammar this way would have been a challenge even to the author himself, let alone a later grammarian.

Globality of evaluation

- Pāṇini's Razor requires in principle that the optimum be recalculated after each new datum.
- This provides at best an idealized model of induction.
- A realistic model will incorporate an inertia factor, which determines the extent to which the order of data presentation and its frequency influences the outcome of acquisition.