How Formal Concept Lattices Solve a Problem of Ancient Linguistics⋆

Wiebke Petersen

Institute of Language and Information
Department of Computational Linguistics
University of Düsseldorf
`petersew@uni-duesseldorf.de`

Abstract. In his grammar of ancient Sanskrit, Pāṇini represents the phonological classes as intervals of a list. This representation method and especially the actual list constructed by Pāṇini, which is called the *Śivasūtras*, earns universal admiration. The legend says that god *Śiva* revealed the *Śivasūtras* to Pāṇini in order to let him start developing his grammar of Sanskrit. A question still discussed is whether it is possible to shorten the *Śivasūtras*. In the course of this paper, I am going to prove that this question can be reduced to a question about the graph-theoretical form of a particular formal concept lattice. Furthermore, I show how the *Śivasūtras* can be reconstructed from Pāṇini's grammar.

1 Introduction

1.1 Pāṇini's Grammar of Sanskrit

Pāṇini's grammar of Sanskrit (see [1], commented edition) is not only one of the oldest recorded grammars (according to [2], it dates from around 350 BC), but also one of the most complete grammars of any language ever written. It earns universal admiration among linguists: "The descriptive grammar of Sanskrit, which Pāṇini brought to its perfection, is one of the greatest monuments of human intelligence and an indispensable model for the description of languages", ([3]). Pāṇini developed a number of ingenious techniques to represent his grammar system in a very compact and concise way, including the introduction of a semi-formalized meta-language and an intricate system of conventions governing rule applications (e.g. [4], [5]). Since the grammar was designed for oral tradition, its compactness was particularly desirable, and the linear form of the whole grammar was a prerequisite.

The science of Sanskrit developed as a tool used for the preservation and propagation of the Vedas, the religious scriptures of ancient Hindus. The various Vedic texts were produced in different regions of India, beginning about 1500 BC (see [6]). In the course of time, a gap developed between the language of

---

the ancient scriptures and the colloquial use of Sanskrit. This gap affected both the phonetic form of the orally preserved texts and the comprehension of them. The late Vedic texts tell an anecdote to illustrate the importance of correct pronunciation and stressing (Śatapatha-Brāhmaṇa 1.6.3.8): The demon Tvaṣṭṛ longed for a son who would kill the war god Indra. But instead of begging for an *indra-śatrú* ('Indra-killer'), he asked for an *índra-śatru* and got a son who was killed by Indra. Hence, a guidance for the correct recitation of the religious texts became necessary. Pāṇini's grammar comprises both the Vedic Sanskrit and bhāṣā, the language spoken by the priestly class of his time, to which we refer as *classical* Sanskrit nowadays. Pāṇini's grammar and its canonization laid the foundations for the development of Sanskrit into a *lingua franca* of adminsitration and science.

Sanskrit is an inflected language with a rich morphology and sandhi.[1] The mastery of the sandhi rules is particularly important since two methods of reciting are used in rituals: the standard method of reciting whole continuous sūtras and the *padapāṭha* – 'word for word'-recitation – of the Veda. For the latter, the sūtras must be analyzed and decomposed into single words and all sandhi-processes must be canceled. This explains why the phonology of Sanskrit stands in the center of interest.

Pāṇini's grammar consists of four components: *Aṣṭādhyāyī*, *Śivasūtras*, *Dhātu-pāṭha*, and *Gaṇapāṭha*. The *Aṣṭādhyāyī* is the central component consisting of about 4.000 rules, which make references to classes defined on the elements of the other three components. The *Śivasūtras* (see Fig. 1) are the smallest component and consist of only 14 sūtras, which comprise a list of the phonological segments of Sanskrit and meta-linguistically used stop markers. According to this list, the natural phonological classes of Sanskrit are defined by a representation method specified in the *Aṣṭādhyāyī*. The *Aṣṭādhyāyī* refer to the phonological classes defined by the *Śivasūtras* in 100s of rules.

अइउण् ॥ऋॡक् ॥एओङ् ॥ऐऔच् ॥हयवरट् ॥लण् ॥अमङणनम् ॥झभञ् ।
॥घढधष् ॥जबगडदश् ॥खफछठथचटतव् ॥कपय् ॥शषसर ॥हल् ॥

$a \cdot i \cdot uṇ \, || \, ṛ \cdot l̥k \, || \, e \cdot oṅ \, || \, ai \cdot auc \, || \, hayavaraṭ \, || \, laṇ \, || \, ñamaṅaṇanam \, || \, jhabhañ \, ||$
$ghaḍhadhaṣ \, || \, jabagaḍadaś \, || \, khaphachaṭhathacaṭatav \, || \, kapay \, || \, śaṣasar \, || \, hal \, ||$

Fig. 1. Pāṇini's *Śivasūtras*

Figure 1 shows the *Śivasūtras* in a linear sequence of sūtras, as constructed by Pāṇini for oral tradition. Nowadays, it is common to present them in the tabular form of Tab. 1 to support readability. Each sūtra consists of a sequence

---

[1] Sandhi refers to the systematic phonological modifications morphemes and words undergo if they are combined. An example of a sandhi phenomenon in English is the variation of the indefinite determiner (a/an), which depends on the first sound of the following word.

Table 1. Pāṇini's *Śivasūtras* in tabular form

1.	a	i	u			Ṇ
2.				ṛ	ḷ	K
3.		e	o			Ṅ
4.		ai	au			C
5.	h	y	v	r		Ṭ
6.					l	Ṇ
7.	ñ	m	ṅ	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

of phonological segments, denoted in the table by lower case letters, followed by one stop marker (called *anubandha*), identified by using capitals. The *anubandhas* are taken from the set of consonants of Sanskrit. As a result, some consonants occur twice in the list: once as an *anubandha* and once as a phonological segment. Furthermore, there is one phonological segment, namely *h*, and one *anubandha*, Ṇ, occurring twice in the same role.

Pāṇini represents the phonological classes of Sanskrit as intervals of the list given by the *Śivasūtras*. Thereby, each class is encoded as a continuous sequence by giving its start segment and the marker element immediately following the last segment of the sequence. Two questions, which are discussed to this day, are whether it is possible to optimize the *Śivasūtras* with respect to the length and how Pāṇini was able to construct the *Śivasūtras*. About the latter, the legend says that god *Śiva* revealed the *Śivasūtras* to Pāṇini in order to let him start developing his grammar of Sanskrit. In the course of this paper, I am going to prove that the question of optimality can be reduced to a question about the graph-theoretical form of a special formal concept lattice. Furthermore, I prove that the *Śivasūtras* can be reconstructed from the *Aṣṭādhyāyī* without making a claim on additional aids. Hence, the hypothesis that the *Śivasūtras* must be necessarily older than the *Aṣṭādhyāyī* proves untenable (e.g. [7]).

We will use a simple phonological rule of Sanskrit as an example to show how the *Śivasūtras* interact with the *Aṣṭādhyāyī*. Phonological rules are operational rules of the form, "*A* is replaced by *B* if preceded by *C* and succeeded by *D*," or in modern notation:[2]

$$A \rightarrow B/_{C\_\_D} \ . \tag{1}$$

As mentioned before, Pāṇini's grammar is designed for oral tradition and hence, non-linguistically signs like '→' cannot be used in rule representations. Pāṇini

---

[2] Alternatively, such a rule can be denoted as a context-sensitive rule: $CAD \rightarrow CBD$.

PROOF

denotes operational rules by using grammatical case markers to prescribe the role an expression plays in a rule. The functions of such meta-linguistically used case suffixes are laid down in individual sūtras of the *Aṣṭādhyāyī*; e.g., the interpretation of the genitive suffix is laid down in sūtra 1.1.49:

1.1.49 *ṣaṣṭhī stāneyogā* (**षष्ठी स्तानेयोगा**)[3] "The function of the genitive case in a sūtra is that of the phrase 'in the place of' when no special rules qualify the sense of the genitive" ([8]).

The four components of (1) are marked by case markers as follows: *A* is marked by the *genitive*, *B* by the *nominative*, *C* by the *ablative*, and *D* by the *locative* suffix.

Sūtra 6.1.77. of the *Aṣṭādhyāyī* serves us as an example for the interaction of the *Śivasūtras* with the *Aṣṭādhyāyī*; it encodes the phonological rule of Sanskrit that the vowels of the class $\langle i, u, ṛ, ḷ \rangle$ are replaced by their non-syllabic (consonantal) counterparts $\langle y, v, r, l \rangle$ if they are followed by a vowel:

6.1.77. *iko yaṇ aci* (**इको यण् अचि**)

Cancelling all sandhi processes results in the *padapāṭha* or 'word-for-word'-form *ikaḥ yaṇ aci*, which is morphologically analyzed as:

$$[\text{ik}]_{\text{genitive}} \quad [\text{yaṇ}]_{\text{nominative}} \quad [\text{ac}]_{\text{locative}} \ .$$

Like the grammatical case markers, the technical expressions **ik**, **yaṇ** and **ac** belong to Pāṇini's meta-language, too; they are called *pratyāhāras* and denote phonological classes. A *pratyāhāra* consists of a phonological segment followed by an *anubandha*. The vowel 'a' in the expression **yaṇ** fulfills two tasks: first, it serves as a linking vowel which turns the *pratyāhāra* into a pronounceable syllable, and second, it prevents the consonant 'y' from being mistaken for the *anubandha* 'Y'. Using the convention of distinguishing the *anubandhas* by capitals, we can write the *pratyāhāras* of sūtra 6.1.77 as **iK**, **yṆ** and **aC**. *Pratyāhāras* denote intervals of the *Śivasūtras*, their interpretation is defined by sūtra 1.1.71 of the *Aṣṭādhyāyī*: A *pratyāhāra* consisting of a phonological marker *a* and an *anubandha M* denotes the continuous sequence of phonological segments in the *Śivasūtras* which starts with *a* and ends with the phonological segment which is the direct predecessor of the *anubandha M*. Table 2 shows the interpretation of the *pratyāhāras* which are involved in sūtra 6.1.77. Note that although Ṇ denotes two distinct *anubandhas*, the meaning of the *pratyāhāra* **yṆ** is unambiguous since only one of the *anubandhas* is a successor of **y**.

Now we are able to state the phonological rule encoded in sūtra 6.1.77 in a modern form:

$$\text{iK} \rightarrow [\text{yṆ}]/\underline{\quad}[\text{aC}] \ .$$

It states that the elements of the class iK= {i,u,ṛ,ḷ} are replaced by their counterparts of the class yṆ= {y, v, r, l} if they occur right in front of a member of the

---

[3] The sūtras are given in Latin transliteration and *Devanāgarī*.

Table 2. Interpretation of the *pratyāhāras* of sūtra 6.1.77. (**iK**= $\{i, u, ṛ, ḷ\}$, **yṆ**= $\{y, v, r, l\}$, **aC**= $\{a, i, u, ṛ, ḷ, e, o, ai, au\}$)

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

class aC= {a,i,u,ṛ,ḷ,e,o,ai,au}. Other sūtras ensure that the correct 'counterpart' is selected.

There are 100s of sūtras in the *Aṣṭādhyāyī* using *pratyāhāras* for the denotation of phonological classes, but altogether not more than 41 different *pratyāhāras* are actually used for this purpose. We will refer to those 41 phonological classes of Sanskrit identified by Pāṇini in the *Aṣṭādhyāyī* as *Pāṇini's phonological classes*.

Phonological segments form a natural phonological class if they behave similarly in similar phonological contexts. This analougous behavior can be expressed in generalized rules as we have seen in sūtra 6.1.77. The phonological classes of a grammar are mutually related: classes can be subclasses of other classes, two or more classes can have common elements, etc. These connections are naturally represented in a hierarchy. The *pratyāhāra* representation encodes such connections in a linear form. An aim of this paper is to determine the conditions under which a set of sets in fact has a *Śivasūtra*-style linear representation.

1.2   The Economy Problem of the *Śivasūtras*

Pāṇini does not discuss the criteria on which he constructed the *Śivasūtras*; but, by looking at the intricate methods he used in the *Aṣṭādhyāyī* to make it as compact as possible, it becomes clear that he aimed at an economical representation.

Since the *Śivasūtras* denote a list of phonological segments and *anubandhas*, two interesting sublists can be regarded: the list of sounds and the *anubandha* list. Hence, the *Śivasūtras* can be optimized in three respects, concerning the length of the lists:

1. The length of the whole list is minimal.
2. The length of the sublist of the *anubandhas* is minimal and the length of the whole list is as short as possible.
3. The length of the sublist of the sounds is minimal and the length of the whole list is as short as possible.

It should be noted that none of these minimality criteria implies one of the others. Looking at the *Śivasūtras*, the double occurrence of the sound h is especially astonishing. This is why the third minimality criterion is in the focus of attention in research on the economy of Pāṇini's *Śivasūtras*. For example, [9] and [10] argue that Pāṇini's *Śivasūtras* are optimal with respect to the third criterion, using linguistic principles and arguments referring to the construction principles of the other parts of the grammar.

By looking at the phonological segments as attributes and at the phonological classes as objects, a formal context can be associated with the *Śivasūtras*.[4] I am going to prove that Pāṇini's *Śivasūtras* respect the third minimality criterion and that this property depends solely on two facts: first, the corresponding formal concept lattice is planar and second, there exists a plane drawing of the Hasse diagram of the lattice in which each attribute concept lies at the boundary of the infinite face if one removes the top node of the lattice and all edges connecting it with co-atoms of the lattice (see Fig. 8).

The rest of this paper is organized as follows: This introductory section will be completed by some preliminary definitions formalizing the third minimality criterion and two short subsections about Formal Concept Analysis and the theory of planar graphs. Section 2 shows that Pāṇini was forced to duplicate at least one of the phonological segments and presents a sufficient condition for the existence of a *Śivasūtra*-style representation. Finally, Sect. 3 explains, how an optimal *Śivasūtra*-style representation can be constructed if it exists. Furthermore, it is proven that solely the duplication of the h enabled Pāṇini to construct the *Śivasūtras* such that they fulfill the third minimality criterion.

1.3   Preliminary Definitions

The following definitions formalize the main concepts of the preceding sections: Definition 1 derives the notion of an *S-alphabet* from the linear form of the *Śivasūtras*. Definition 2 and Def. 3 generalize Pāṇini's method of using *pratyāhāras* to represent phonological classes. The phenomenon of the duplicated phonological segment h in Pāṇini's *Śivasūtras* is covered by the notion of an *enlarged S-alphabet*. Finally, Def. 4 formalizes the third minimality criterion.

Definition 1. *A* well-formed Śivasūtra-alphabet *(short* S-alphabet*) is a triple* $(\mathcal{A}, \Sigma, <)$ *consisting of two disjoint finite sets* $\mathcal{A}$ *and* Σ*, and a total order* $<$ *on* $\mathcal{A} \cup \Sigma$*.* $\mathcal{A}$ *is called the* alphabet *and* Σ *the* marker set*.*

Definition 2. *A subset* T *of the alphabet* $\mathcal{A}$ *is* S-encodable *in* $(\mathcal{A}, \Sigma, <)$ *if and only if there exists* $a \in \mathcal{A}$ *and* $M \in \Sigma$*, such that* $T = \{b \in \mathcal{A} | a \leq b < M\}$*.* aM *is called the* pratyāhāra *or* S-encoding *of* T *in* $(\mathcal{A}, \Sigma, <)$*.*

In the following, we call a pair $(\mathcal{A}, \Phi)$ consisting of a finite set $\mathcal{A}$ and a set Φ of subsets of $\mathcal{A}$ (i.e., $\Phi \subseteq \mathbf{P}(\mathcal{A})$) a *system of sets*.

---

[4] Formal contexts and formal concept lattices are defined in Sec. 1.4.

Definition 3. *An S-alphabet $(\mathcal{A}', \Sigma, <)$ corresponds to a system of sets $(\mathcal{A}, \Phi)$ if and only if $\mathcal{A} = \mathcal{A}'$ and each element of Φ is S-encodable in $(\mathcal{A}', \Sigma, <)$. An S-alphabet which corresponds to $(\mathcal{A}, \Phi)$ is called an* S-alphabet of $(\mathcal{A}, \Phi)$. *A system of sets for which a corresponding S-alphabet exists is said to be* S-encodable.

For example, take the set of subsets

$$\Phi = \{\{d, e\}, \{b, c, d, f, g, h, i\}, \{a, b\}, \{f, i\}, \{c, d, e, f, g, h, i\}, \{g, h\}\} \quad (2)$$

of the alphabet $\mathcal{A} = \{a, b, c, d, e, f, g, h, i\}$; it is S-encodable and

$$a\, b\, M_1\, c\, g\, h\, M_2\, f\, i\, M_3\, d\, M_4\, e\, M_5 \quad (3)$$

is one of the corresponding S-alphabets. The S-encodings of Φ are: dM_5, bM_4, aM_1, fM_3, cM_5, and gM_2.

In order to formalize the third minimality criterion and to deal with the double occurrence of h in the Śivasūtras, we need the concept of *enlarging* an S-alphabet: $\hat{\mathcal{A}}$ is said to be an *enlarged alphabet* of $\mathcal{A}$ if there exists a surjective map $\vartheta : \hat{\mathcal{A}} \to \mathcal{A}$. ϑ extends naturally to sets: $\vartheta : \mathbf{P}(\hat{\mathcal{A}}) \to \mathbf{P}(\mathcal{A})$. It is clear that for every system of sets $(\mathcal{A}, \Phi)$ we can find an enlarged alphabet $\hat{\mathcal{A}}$ and a set of subsets $\hat{\Phi}$ with $\Phi = \{\vartheta(\varphi') : \varphi' \in \hat{\Phi}\}$ such that $(\hat{\mathcal{A}}, \hat{\Phi})$ is S-encodable. To achieve such an S-encodable system of sets $(\hat{\mathcal{A}}, \hat{\Phi})$ we enlarge $\mathcal{A}$ so that the sets of $\hat{\Phi}$ are pairwise disjoint. Then we arrange the sets of $\hat{\Phi}$ in an arbitrary sequence and separate them by markers. The induced S-alphabet $(\hat{\mathcal{A}}, \hat{\Sigma}, \hat{<})$ corresponds obviously to $(\hat{\mathcal{A}}, \hat{\Phi})$.

An S-alphabet of $(\hat{\mathcal{A}}, \hat{\Phi})$ will sometimes be called an *enlarged S-alphabet of* $(\mathcal{A}, \Phi)$. Since we always find a finite, enlarged S-alphabet of $(\mathcal{A}, \Phi)$, a minimally enlarged S-alphabet exists.

Definition 4. *An enlarged S-alphabet $(\hat{\mathcal{A}}, \hat{\Sigma}, \hat{<})$ of $(\mathcal{A}, \Phi)$ is said to be* optimal *if and only if it fulfills the following conditions: First, there exists no other enlarged S-alphabet $(\tilde{\mathcal{A}}, \tilde{\Sigma}, \tilde{<})$ of $(\mathcal{A}, \Phi)$, the alphabet $\tilde{\mathcal{A}}$ of which has fewer elements than $\hat{\mathcal{A}}$ and furthermore, as a secondary condition, no other enlarged S-alphabet $(\tilde{\mathcal{A}}, \tilde{\Sigma}, \tilde{<})$ of $(\mathcal{A}, \Phi)$ exists with $|\tilde{\mathcal{A}}| = |\hat{\mathcal{A}}|$ and $|\tilde{\Sigma}| < |\hat{\Sigma}|$.*

1.4   Formal Concept Analysis

Formal Concept Analysis (see [11]) starts with the definition of a *formal context* K as a triple (G, M, I) consisting of a set of objects G, a set of attributes M, and a binary *incidence relation* $I \subseteq G \times M$. For any subset of objects $A \subseteq G$, their set of common attributes is defined as $A' := \{m \in M | \forall g \in A : (g, m) \in I\}$. Analogously, the set of common objects for $B \subseteq M$ is $B' := \{g \in G | \forall m \in B : (g, m) \in I\}$. A *formal concept* is a pair (A, B) with the properties $A = B'$ and $B = A'$, where A is called the *extent* and B the *intent* of the concept. The set of all formal concepts of a context is partially ordered by the subconcept-superconcept-relation: $(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow B_1 \supseteq B_2$. The set of formal concepts together with this partial order forms a complete lattice, called

the *formal concept lattice*. As usual, we denote the set of all formal concepts of a formal context (G, M, I) by $\mathcal{B}(G, M, I)$ and the corresponding concept lattice by $\underline{\mathcal{B}}(G, M, I)$. The *attribute concept* $\mu(m)$ associated with an attribute m is the greatest concept whose intent contains m, and analogously, the *object concept* $\gamma(g)$ of an object g is the smallest concept whose extent contains g.

Formal Concept Analysis has been applied to a number of linguistic problems before. A survey of these linguistic applications can be found in [12]. What is remarkable about the application discussed in the present paper is that the attention is focused on the graph-theoretical form rather than the order-theoretical form of the formal concept lattices as it is generally the case.

1.5   Criterion of Kuratowski on Planar Graphs

If a graph can be drawn in the Euclidean plane, such that neither a vertex nor a point of an edge lies in the inner part of another edge (i.e., no crossing edges exist), then it is said to be *planar*.[5] A lattice is said to be a *planar lattice* if its Hasse-diagram augmented by an extra edge from the top to the bottom node is a planar graph. One of the most important criteria for the planarity of graphs is the criterion of Kuratowski, which is based on the notion of minors of a graph. A graph M is said to be a *minor* of a graph G if it can be arrived from G by first removing a number of vertices and edges from G and then contracting some of the remaining edges.

Proposition 1 (Criterion of Kuratowski). *A graph G is planar if and only if G contains neither a K^5 nor a $K_{3,3}$ as a minor (see Fig. 2).*
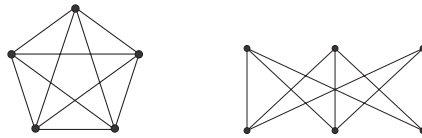
Fig. 2. The complete graph K^5 with 5 vertices (left) and the complete bipartite graph $K_{3,3}$ with $2 \cdot 3$ vertices (right)

2   Planar Formal Concept Lattices and S-encodability

2.1   Are Pāṇini's Phonological Classes of Sanskrit S-encodable?

To each system of sets $(\mathcal{A}, \Phi)$ we define the *corresponding context* $(\Phi, \mathcal{A}, \ni)$. The Hasse-diagram of $\underline{\mathcal{B}}(\Phi, \mathcal{A}, \ni)$ gives us a first hint on whether $(\mathcal{A}, \Phi)$ is S-encodable:

---

[5] Formal definitions of planar graphs and planarity can be found in most textbooks on graph theory (e.g. [13]).

Proposition 2. *If $(\mathcal{A}, \Phi)$ is S-encodable, then the corresponding formal concept lattice $\underline{\mathcal{B}}(\Phi, \mathcal{A}, \ni)$ is planar.*

Proof. Let $(\mathcal{A}, \Sigma, <)$ be an S-alphabet of $(\mathcal{A}, \Phi)$. The proof is based on the construction of a plane drawing of the Hasse-diagram of $\underline{\mathcal{B}}(\mathcal{A}, \Phi, \in)$ with stair-shaped edges.[6] For each formal concept (A, B) of $(\mathcal{A}, \Phi, \in)$, its coordinates in $\mathbf{R}^2$ are given as follows: The smallest object of A w.r.t. $(\mathcal{A}, \Sigma, <)$ determines the x-coordinate of the vertex;[7] its y-coordinate is given by the length of the longest descending chain between (A, B) and $(\emptyset'', \emptyset')$ in $\underline{\mathcal{B}}(\mathcal{A}, \Phi, \in)$.

The edges of the constructed Hasse-diagram are stair-shaped polygonal arcs: Let (A, B) and $(\bar{A}, \bar{B})$ be two formal concepts of $(\mathcal{A}, \Phi, \in)$ with $(A, B) \prec (\bar{A}, \bar{B})$. If $\min(A) = \min(\bar{A})$, then the edge between (A, B) and $(\bar{A}, \bar{B})$ is a straight line; else the vertices (A, B) and $(\bar{A}, \bar{B})$ are connected by the polygonal arc (see Fig. 3)

$$((\bar{A}, \bar{B})_x, (\bar{A}, \bar{B})_y), ((\bar{A}, \bar{B})_x, (\bar{A}, \bar{B})_y - \frac{1}{2}),$$

$$((A, B)_x - \frac{1}{2}, (\bar{A}, \bar{B})_y - \frac{1}{2}), ((A, B)_x - \frac{1}{2}, (A, B)_y), ((A, B)_x, (A, B)_y) \ .$$

The only exception to this edge-construction rule is that every edge between a concept (A, B) and $(\emptyset'', \emptyset')$ is just a straight line. The construction of the edges guarantees that no vertex of the Hasse-diagram lies in the inner part of an edge.
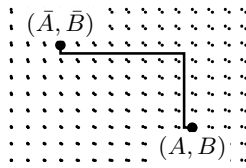
Fig. 3. Stair-shaped edge of the Hasse-diagram

With a simple but detailed case distinction it can be proven that every conflict which occurs between two edges (i.e. every crossing of two edges) can be solved by slightly transforming one of the edges in such a way that the distance between the transformed and the original edge does not exceed $\frac{1}{4}$. □

Figure 4 shows the resulting plane Hasse-diagram of $\underline{\mathcal{B}}(\mathcal{A}, \Phi, \in)$ with Φ and $\mathcal{A}$ taken from example (2); the vertices are placed w.r.t. the S-alphabet given in (3).

It follows as a corollary that a system of sets is not S-encodable whenever the corresponding formal concept lattice is not planar.

Together with Kuratowski's criterion this proves that Pāṇini's phonological classes are not S-encodable since Fig. 5 shows a section of the concept lattice

---

[6] Rotating the constructed Hasse-diagram by $180°$ yields in a plane Hasse-diagram of $\underline{\mathcal{B}}(\Phi, \mathcal{A}, \ni)$.

[7] If the smallest element of A is the n-th smallest element of $\mathcal{A}$ in $(\mathcal{A}, \Sigma, <)$, then the x-coordinate is n. If $A = \emptyset$, then the x-coordinate is 0.
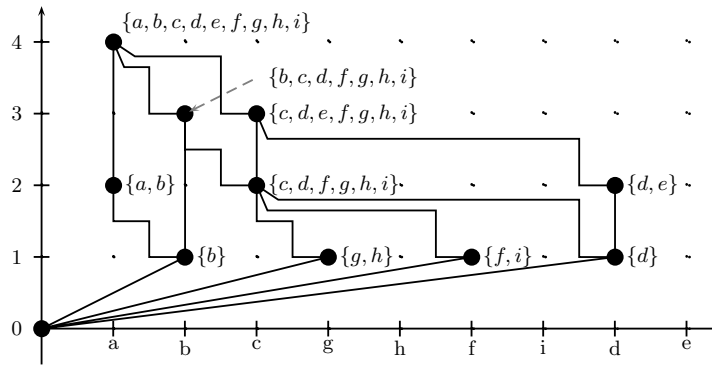
Fig. 4. Stair-shaped plane Hasse-diagram of the concept lattice of $(\mathcal{A}, \Phi, \in)$ with Φ and $\mathcal{A}$ taken from (2); the vertices are placed w.r.t. the S-alphabet in (3).

corresponding to the phonological classes, which has K^5 as a minor. Hence, Pāṇini was forced to duplicate at least one of the phonological segments. But it remains to prove that h is the best candidate for the duplication; this discussion will be postponed.

Proposition 3. *Pāṇini's phonological classes of Sanskrit are not S-encodable.*

2.2   Excursus: A Sufficient Condition of S-encodability

The condition for S-encodable systems of sets given in Prop. 2 is necessary but not sufficient, however. Figure 6 (left) shows an example of a system of sets which is not S-encodable, although its corresponding concept lattice is planar. We need a stronger condition to fully identify those systems of sets which are S-encodable.

Proposition 4. *Let $(\mathcal{A}, \Phi)$ be a system of sets and $\bar{\Phi} = \Phi \cup \{\{a\} : a \in \mathcal{A}\}$. The following statements are equivalent:*
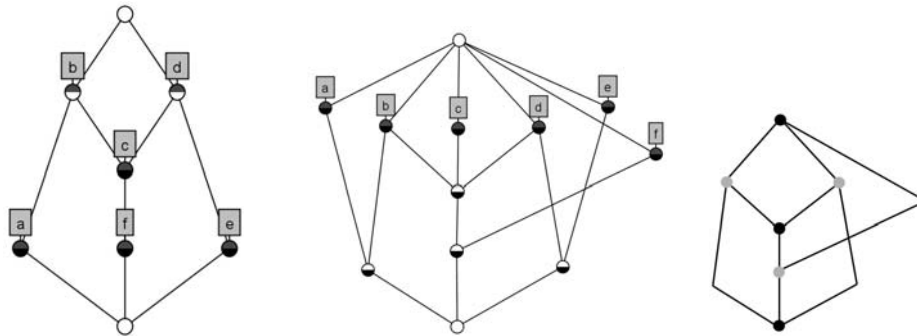
1. *$(\mathcal{A}, \Phi)$ is S-encodable.*
2. *$\underline{\mathcal{B}}(\bar{\Phi}, \mathcal{A}, \ni)$ is planar.*

Proof. By adding a new singleton $\{a\}$, $a \in \mathcal{A}$, to Φ, the S-encodability is preserved (at most one new marker immediately following a has to be inserted in an S-alphabet of $(\mathcal{A}, \Phi)$). Hence, $(\mathcal{A}, \Phi)$ is S-encodable if and only if $(\mathcal{A}, \bar{\Phi})$ is S-encodable. Together with Prop. 2, this proves that statement 1. implies statement 2.

Given a plane drawing of $\underline{\mathcal{B}}(\bar{\Phi}, \mathcal{A}, \ni)$, all singletons of $\bar{\Phi}$ are co-atoms of the lattice and the x-coordinates of their corresponding attribute concepts induce a total order on $\mathcal{A}$. Inserting a marker behind each element of $\mathcal{A}$ yields an S-alphabet encoding of $(\mathcal{A}, \bar{\Phi})$. □
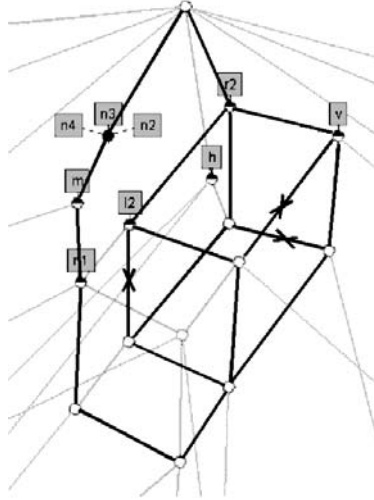
Fig. 5. The figure shows a section of the concept lattice corresponding to Pāṇini's phonological classes which has K^5 as a minor. The minor K^5 can be derived by deleting all but the highlighted edges and contracting the edges marked by crosses. The figure shows that the class memberships of the phonological segments h, v and l (denoted by $l2$) are independent of each other.

Fig. 6. Left: formal concept lattice of $(\Phi, \mathcal{A}, \ni)$ with $\Phi = \{\{d, e\}, \{a, b\}, \{b, c, d\}, \{b, c, d, f\}, \{a, b, c, d, e, f\}\}$; middle: formal concept lattice of $(\bar{\Phi}, \mathcal{A}, \ni)$ which has $K_{3,3}$ as a minor (see right figure)

It follows as a corollary, that a system of sets $(\mathcal{A}, \Phi)$ is S-encodable whenever a plane Hasse-diagram of $\underline{\mathcal{B}}(\Phi, \mathcal{A}, \ni)$ exists in which each attribute concept lies at the boundary of the infinite face if one removes the vertex $(\emptyset', \emptyset'')$ from the Hasse-diagram. This boundary graph is called the *S-graph of* $(\mathcal{A}, \Phi)$ and it is fixed up to isomorphism. The left part of Fig. 7 shows a plane drawing of the formal concept lattice corresponding to (2) in which the S-graph is highlighted.

Looking back at the example given in Fig. 6, it is clear that by moving from Φ to $\bar{\Phi}$ the concept lattice loses the quality of being planar; $\underline{\mathcal{B}}(\bar{\Phi}, \mathcal{A}, \ni)$ has the bipartite graph $K_{3,3}$ as a minor (figure on the right side). Hence, $(\mathcal{A}, \Phi)$ cannot be S-encodable; this can be also derived from the fact that the attribute concept $\mu(f)$ does not lay in the S-graph of the Hasse diagram of $\underline{\mathcal{B}}(\Phi, \mathcal{A}, \ni)$ (left figure).

3   Constructing *Śivasūtra*-Style Representations

### 3.1   The S-graph Determines the S-alphabet

If $(\mathcal{A}, \Phi)$ is a system of sets which is S-encodable, then an S-alphabet $(\mathcal{A}, \Sigma, <)$ of $(\mathcal{A}, \Phi)$ can be found as follows: Take the labeled S-graph of $(\mathcal{A}, \Phi)$ and a path in it, that starts and ends at the vertex corresponding to $(\mathcal{A}', \mathcal{A}'')$. The path must meet the following conditions: First, the path passes each attribute concept at least once; second, none of the edges occurring more than once in the path is part of a cycle in the S-graph. By looking at the S-graph as a subgraph of the directed Hasse-diagram, the edges of the path can be directed.

The S-alphabet, seen as a sequence of markers and elements of $\mathcal{A}$, can be constructed from the empty sequence by traversing the path from the beginning to the end: If an attribute concept $\mu(a)$ is reached, then add a to the sequence. If an edge is passed whose direction contradicts the traversal direction, a new, previously unused, marker element is added to the sequence, unless the last added element is already a marker. Finally, after the end of the path is reached, revise the sequence as follows: If an element of $\mathcal{A}$ appears more than once in the sequence, delete all but the first occurrences. The definition of the S-graph guarantees that, if the path passes an attribute concept $\mu(a)$ more than once, the path goes upwards immediately after it reaches $\mu(a)$ for the first time. Hence, eliminating all but the first occurrence of a reduces the number of markers in the resulting S-alphabet.

Applied to our small example (2), we may choose the path illustrated in Fig. 7, which fulfills the required conditions. Traversing the path, we pass first $\mu(a)$ and $\mu(b)$ without using an edge against its destined direction. Now we move downwards and violate the direction of the edge, and therefore we have to add a marker to our sequence, so that it starts with $a\,b\,M_1$. Now moving upwards we collect the c and the g, but since $\mu(g) = \mu(h)$ we also have to collect the h. After this we move downwards again, and that is why we add a new marker. We again reach $\mu(c)$ and add c a second time to our sequence. So far our sequence is $a\,b\,M_1\,c\,g\,h\,M_2\,c$, and if we continue we end up with the S-alphabet depicted in (3).

Note that this procedure does not yield a unique S-alphabet since we have several decisions to make: (a) If two attribute concepts are identical the order of the attributes in the S-alphabet is arbitrary; (b) from $\mu(c)$ we can either go to $\mu(g)$ or $\mu(i)$; (c) the path can be traversed clockwise or anti-clockwise.

Whenever a run violates the direction of an edge immediately after passing an attribute concept, a new marker has to be added to the S-alphabet. Hence,
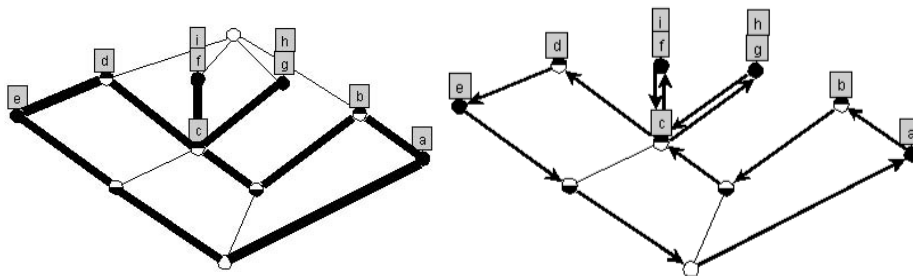
Fig. 7. Left: plane Hasse-diagram of the formal concept lattice corresponding to (2) with highlighted S-graph; right: possible path in the S-graph from which the S-alphabet $a\,b\,M_1\,c\,g\,h\,M_2\,f\,i\,M_3\,d\,M_4\,e\,M_5$ can be achieved

every optimal S-alphabet of $(\mathcal{A}, \Phi)$ can be constructed by finding a run through the S-graph which minimizes the number of such marker-insertion situations.

3.2 Pāṇini's *Śivasūtras* Are Optimal

Figure 8 shows a plane Hasse-diagram of the concept lattice of $(\hat{\Phi}, \hat{\mathcal{A}}, \ni)$, where $\mathcal{A}$ is the alphabet of phonological segments of Sanskrit, Φ is the set of Pāṇini's phonological classes, and $\mathcal{A}$ and Φ are enlarged by duplicating the segment h according to Pāṇini (the duplication of h is denoted by $h\_$).[8] The black and the striped rectangles next to some of the vertices mark the places where markers have to be added, depending on the traversal direction (black: anti-clockwise [14 markers], striped: clockwise [17 markers]). It is obvious that no S-encoding can have less than 14 markers and the optimal S-alphabets are the various combinatorial variants of

$\langle$a,i,u,M_1,ṛ,l,M_2, $\{\langle\{$e,o$\}, M_3\rangle, \langle\{ai,au\}, M_4\rangle\}$,h,y,v,r,$M_5$,l,$M_6$,

ñ,m,$\{$n̄,ṇ,n$\}$, M_7, jh,bh,M_8, $\{$gh,ḍh,dh$\}$, M_9,j,$\{$b,g,ḍ,d$\}$, M_{10},

$\{$kh,ph$\}$, $\{$ch,ṭh,th$\}$, $\{$c,ṭ,t$\}$, M_{11}, $\{$k,p$\}$, M_{12}, $\{$ś,ṣ,s$\}$, M_{13},h,$M_{14}\rangle$.

among which Pāṇini's *Śivasūtras* can be found. [9] argues in detail that the order chosen by Pāṇini out of the set of possibilities is unique if one requires a subsidiary principle of restrictiveness.

So far we have argued that Pāṇini was forced to enlarge the alphabet of phonological segments, but it remains to show why duplicating the h is the best choice. If h is entirely removed from the 41 phonological classes, then the optimal S-alphabet has only one marker less, namely 13.

In Pāṇini's phonological classes the phonological segments h, v, and l occur independently of each other. The Hasse-diagram of a concept lattice of a formal context which contains three independent attributes has K^5 as a minor and is therefore not planar (see Fig. 5). Triples of three independent attributes are called K^5-*triples*. Hence, to get a planar concept lattice it is necessary to duplicate at least one element of each K^5-triple.

---

[8] Drawings done by 'Concept Explorer' (http://www.sourceforge.net/projects/conexp).
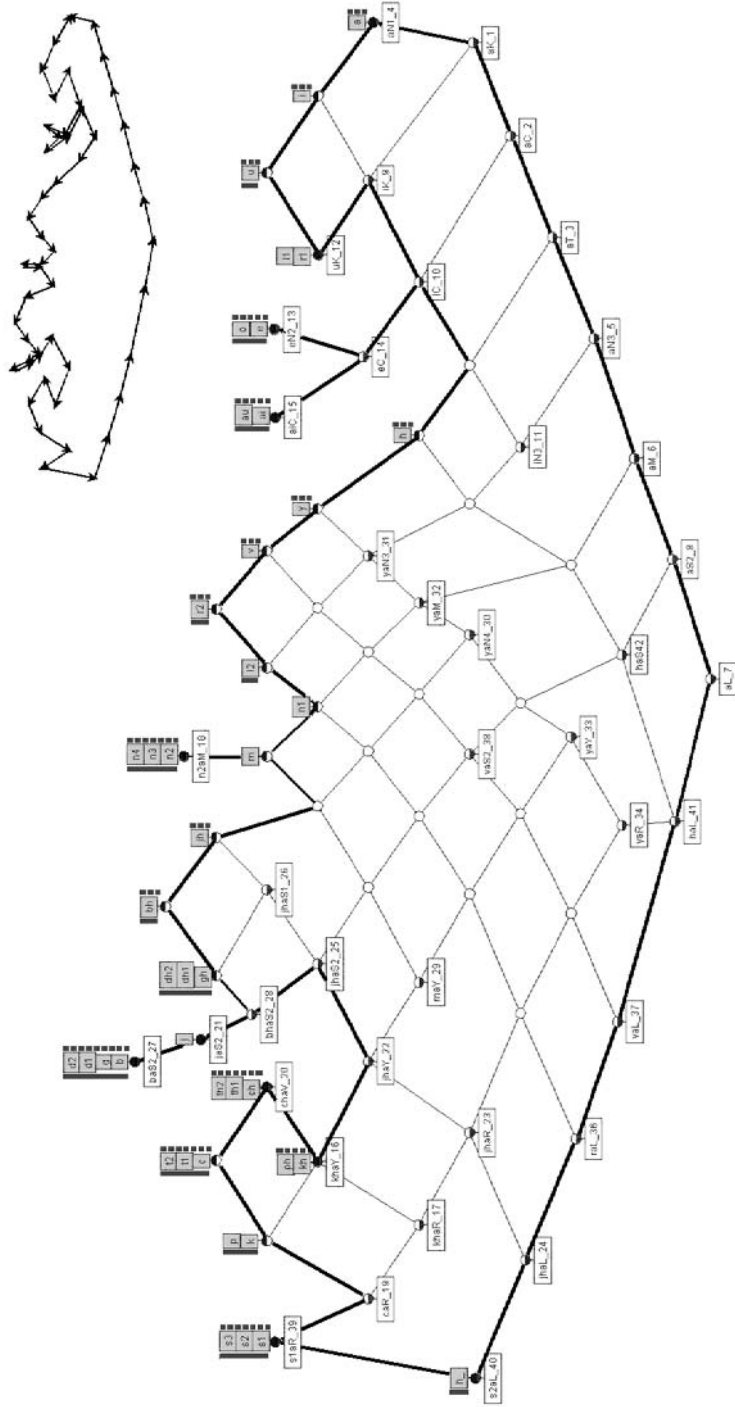
Fig. 8. Plane Hasse diagram of the concept lattice of Pāṇini's phonological classes of Sanskrit. The denotations in the figure are as follows: h is the duplicate of h and r1:r, l1:l, r2:r, l2:l, n1:ñ, n2:ṅ, n3:ṇ, n4:n, dh1:ḍh, dh2:dh, d1:ḍ, d2:d, th1:ṭh, th2:th, t1:ṭ, t2:t, s1:ś, s2:ṣ, s3:s.

The white boxes mark the phonological classes. The small figure on top shows the path in the S-graph we have to choose in order to construct Pāṇini's *Śivasūtras*.

Looking at Pāṇini's phonological classes, we find 249 K^5-triples; each of them contains h, and no other element is contained in each of them. Hence, to avoid the duplication of h it would be necessary to duplicate more than one element. For this reason, there is no other choice then duplicating h in order to get an optimal S-alphabet corresponding to Pāṇini's phonological classes.

This answers the question whether Pāṇini's *Śivasūtras* are optimal in the sense that there exists no other sequence of the phonological segments interrupted by less stop markers.

Proposition 5. *Pāṇini's* Śivasūtras *form an optimal S-alphabet.*

4   Outlook

Pāṇini's method of linearly encoding a subset of a power set could also be interesting from the viewpoint of other coding and sorting problems. However, it should be noted that the property of S-encodability with no or only moderate enlargement of the alphabet set is rare, at least among the phonological systems of natural languages (see [14]). Hence, linguists should investigate how Pāṇini's phonological analysis of Sanskrit differs from phonological classifications of other languages. Furthermore, it would be interesting to give an alternative mathematical description of S-encodable systems of sets and enlarged S-alphabets for testing and using the property.

References

1. Katre, S.M.: Aṣṭādhyāyī of Pāṇini. University of Texas Press, Austin (1987)
2. Kiparsky, P.: Paninian linguistics. In Asher, R.E., ed.: The Encyclopedia of Language and Linguistics. Volume 6. Pergamon Press, Oxford (1994)
3. Bloomfield, L.: Review of Liebich, Konkordanz Panini-Candra. Language **5** (1929) 267–276
4. Kiparsky, P.: On the architecture of Pāṇini's grammar. Lecture notes (2002)
5. Ostler, N.: Sanskrit studies as a foundation for computational linguistics. In: Proceedings of the LESAL Workshop, Mumbai (2001)
6. Deshpande, M.M.: Ancient Indian phonetics. In Koerner, E.F.K., Asher, R.E., eds.: Concise history of the language sciences: From the Sumerians to the cognitivists. Elsevier, Oxford, New York, Tokyo (1995) 72–77
7. Faddegon, B.: The mnemotechnics of Pāṇini's grammar I: The Śiva-Sūtra (1929). In Staal, F., ed.: A Reader on the Sanskrit Grammarians. Motilal Banarsidass (1985) 275–285
8. Vasu, S.C., ed.: The Aṣṭādhyāyī of Pāṇini, Allahabad (1891) (2 volumes). Reprint: Delhi (1962)
9. Kiparsky, P.: Economy and the construction of the Sivasutras. In Deshpande, M.M., Bhate, S., eds.: Paninian Studies. Ann Arbor, Michigan (1991)
10. Staal, F.J.: A method of linguistic description. Language **38** (1962) 1–10
11. Ganter, B., Wille, R.: Formal Concept Analysis. Mathematical Foundations. Springer, Berlin (1999)

12. Priss, U.: Linguistic applications of Formal Concept Analysis. (In: Proceedings of ICFCA 2003)
13. Diestel, R.: Graph Theory. Springer, New York (1997)
14. Petersen, W.: A mathematical analysis of Pāṇini's Śivasūtras. Journal of Logic, Language and Information **13** (2004) 471–489