Šefčík, Ondřej

**Strings, oppositions, metrics**

*Linguistica Brunensia.* 2009, vol. 57, iss. 1-2, pp. [7]-17

ISBN 978-80-210-4891-1
ISSN 1803-7410 (print); ISSN 2336-4440 (online)

LINGUISTICA BRUNENSIA 57, 2009, 1–2

## STATI – СТАТЬИ – ARTICLES – AUFSÄTZE

ONDŘEJ ŠEFČÍK

# STRINGS, OPPOSITIONS, METRICS

(This paper was written with the help of Czech Science Foundation,
# 405/06/P179)

## 0 Preliminaries

Any phonological theory deals with phonemic elements, strings of phonemic elements and oppositions between phonemic elements and strings.

In the present paper we will focus more on the way the set of phonemic elements behind the set of phonemic strings is identified.

By the term **phonemic element** is meant such a phonemic segment which is successfully testified as a minimal element, i.e. not a union of more segments of given segmentation (including the well-known commutation test). So in the list: such elements could be phones, phonemes, intonation, stress or juncture (diaereme). Any such element can be localized in the string of other elements *sui generis* and distinguished from them. Such a procedure will be demonstrated below.

**Note 1:** In the following lines it is only phonemes that are considered; however, common procedures described below could be applied on other sets of phonemic elements, too.

**Note 2:** All data are excerpted from the Old Indo-Aryan (Vedic) language. Data are present in the canonical form, not in any kind of phonetic or phonemic transcription. For the sake of simplicity we offer only such examples in which phonemes can be easily, if necessary, described.

A **phonemic opposition** is a relation of distinction between two or more elements of the same set. Sometimes a phonemic opposition is defined in terms of traditional structural linguistics as a paradigmatic relation; both definitions are not mutually exclusive.

In the following lines we will show that not only sets themselves but even strings can be compared with the use of the theory of oppositions and that any opposition can be identified only by the means of a comparison of two or more strings of a given language.

In what follows we will try to sketch an analytical method of identification of oppositions on strings using metric spaces.

Note 3: From here on we omit the adjective phonemic, as we speak only about phonemic oppositions. However, most of the theses postulated here can be applied on phones, too.

## 1 Strings and their properties

Let us suppose a set of phonemic elements *P*. Any finite sequence of elements of the set *P* (elements can be repeatedly present in sequence) will be called a **string**, symbolically marked by σ. A void string is taken as a string *sui generis*, too.

**Note 4:** If we set out from the set of phonemic elements and we consider the operation of concatenation, then we have an algebra called *semigroup.* A semigroup which contains also a null string is *monoid*. The operation of concatenation makes strings from the elements of the set of phonemic elements. But not all possible strings are existing strings of a given language. Such existing strings are proper subsets of the set of possible strings. However, if following the analytical procedure, we face up, from the beginning, only existing strings (see Brainerd 1971: 109–111).

Strings interesting for phonologists are words, morphs and morphemes. An analysis of words into morphs and morphemes is generally known (and such an analysis does not differ in principles from an analysis of morphs and morphemes into phones and phonemes as described below).

The minimal morphemic string is such a substring of any other string which is the minimal union of linguistic form and meaning in the sense of de Saussure and his followers, i.e. it cannot be divided into any such strings or in other words, the number of possible subsets of the same degree of analysis is equal to one.

In the rest of this section we will talk about properties of strings from the general perspective. All properties given below are valid not only in general but could also be applied on strings of phonemes. In the examples which follow the first of each pair is generally oriented, the second example is based on the OIA material.

## 2 Properties of strings of elements

Every string has its length given by the number of elements the string contains. A void string has the length equal to zero.

**Example 1:** The string *ba* has a different number of elements in contrast with the string *bac* or the string *b*. The first string has the length equal to *2*, the second string equal to *3*, the third to *1*.
**Example 2:** The string *asti* has a different number of elements in contrast with the string *asi* (however, both consist of two morphemes, the first of them being identical). The first string has the length equal to *4*, the second equal to *3*.

In any string there can hence be present, first, the term of distance between elements and second, **linear metrics** over tuple (*P, l*) where *P* is a set of elements of the string and *l* is linear distance between such elements, which was described in the paper (Šefčík 2007: 19–22).

We should keep in mind that any string is an **ordered set** of elements. Ordering simply means that it does matter in which position an element occurs in a given string.

**Example 3:** Leaving out the ordering of a string, two strings *man* and *nam* are equal in the number and values of their elements, hence both strings are in null opposition. So are the strings *mna, anm, amn, nma* equal to them. Such a statement is contradiction with the common usage of language the position and thus ordering matters.
**Example 4:** In accordance with the preceding example, the OIA strings *manyate* and *namyate* are two different strings (words) with different meanings.

From the above-written example we see that the ordering of elements in the set is important for linguistic analysis just as the length of a string is.

Comparing two strings of the same length we can see that strings can differ not only in the length of a string, but in elements, too.

**Example 5:** The string *ba* differs from the string *ca* in elements; however, both strings have the same length equal to *2*.
**Example 6:** The strings *vardhati* and *varṣati* have the same length, they differ only in one element. Other elements are in both strings in the same order.

The differences between strings of the same length could be expressed with the use of **gross metrics** which expresses the distance between strings by counting the number of distinctions of elements in the same positions of compared strings. A more precise definition of gross metrics was given in (Šefčík 2007: 22–23); additional information will be given bellow.

If there is no distinction between elements in the same position in a string of the same length, we will mark such position with *0*. As a tool for describing a distance between strings we can employ the Hamming distance.

The Hamming distance is the number of positions for which the corresponding symbols are different. Put another way, it measures the minimum number of *substitutions* required to change one into the other, or the number of *errors* that transformed one string into the other (Hamming 1950: 147–160).

Gross metrics is then a tuple ($P$, $g$) where $P$ is a set of elements and $g$ is a Hamming distance between them.

It is gross metrics which expresses the state of oppositions between any elements.

For phonologists, the most important is to find such strings which are in mutual distance equal to *1*, i.e. between strings of the same length, with a difference only in one element. Such strings are in **minimal opposition** then, or in other words, between pair of elements there is an opposition only in case there is at least one pair of strings differing only in the presence of one element in the first string and the presence of the second element in the second string. Any strings which are not in minimal opposition we suppose to constitute a bundle of such minimal oppositions; we take for granted that any such a bundle could be partitioned into such minimal oppositions. The distance in gross metric space between any strings can be expressed with the use of the sum of differences. The $\Sigma = 0$ means that there is no opposition between strings (this example is trivial but must be mentioned here), the $\Sigma = 1$ means that there is a minimal opposition, the $\Sigma < 1$ expresses a non-minimal oppositions, the number then means the number of minimal oppositions which this non-minimal opposition contains.

This process of stating minimal oppositions between strings is generally known as a commutation or substitution test, which is widely used since the beginning of modern phonology. Again, we should keep in mind that the substitution test is applied on whole strings.

By the partitioning of elements, which make the distinction between those strings, we find elements which form the strings. And indeed – if a supposed element, which makes the distinction between strings, is really such an element of one of the strings, it is not an element of another string.

Correctly done substitution leads us towards the finite set of elements of strings.

**Note 5:** All preceding and following statements on strings can be applied on any substring of any string, too.
**Note 6:** From the above-written it is clear that any phonological theory begins with the analysis of morphs and morphemes according to the hierarchy of speech analysis.

**Example 7:** The strings *abc* and *abd* are in a minimal opposition, expressed with the use of the Hamming distance:

| | |
|---|---|
| *abc* | |
| *abd* | |
| 001 | $\Sigma = 1$ |

**Example 8:** Vedic *dadāti* and *dadhāti* are in a minimal opposition, which can be expressed in the following table:

| | |
|---|---|
| *dadāti* | |
| *dadhāti* | |
| 001000 | Σ =1 |

**Example 9:** The strings *abc* and *ade* are in a non-minimal opposition, being the sum of two minimal oppositions:

| | |
|---|---|
| *abc* | |
| *ade* | |
| 011 | Σ = 2 |

**Example 10:** Vedic *kalpate* and *kramate* are in a non-minimal opposition, containing three minimal oppositions:

| | |
|---|---|
| *kalpate* | |
| *kramate* | |
| 0111000 | Σ =3 |

If in the preceding line the substitution was carried out correctly with the help of gross metrics, then we face an unordered set of elements of strings.

Such a finite set is unordered in the sense that as the lack of ordering we interpret the lack of a system. As a system we understood a set composed not only of elements, but also of relations between the elements. So it is the organization inside the set which make the ordering of a set, i.e. the system.

## 3 Oppositions between strings

A relation between strings is an opposition between strings which could be classified as any other opposition with the use of terms **base** and **complement** of opposition. We defined both terms on examples of opposition between values of phonemes in our paper (Šefčík 2008a: 11–13). Here we will reformulate our previous statement which is based on ideas by Marcus and Brainerd (Marcus 1967: 4–26, 79–83, Marcus 1969: 15–33, 77–80, Brainerd 1971: 20–22) who reformulated ideas by Trubetzkoy (1939: 59–79) and Cantineau (1952: 11–40, 1955: 1–9).

The base of an opposition between strings is such elements of the string which are identical for all considered strings. Such elements are expressed, with gross metrics used, in a Hamming code as *0*s. Or in other words, the base is the intersection of given strings, the common part of those strings.

On the contrary, the complement of an opposition is such elements which are not identical for all considered strings. Such elements are expressed, gross metrics used again, in a Hamming code as *1*s. In other words: the complement is that part of values of strings which is not common to all the strings considered.

**Example 11:** If we return to an example mentioned above with the strings *abc* and *abd* in a minimal opposition, we can say that the base is the substring *ab*, the substrings (here equal to sole elements) *c* and *d* being complements of this opposition.

**Example 12:** Again, let us return to an example mentioned above, the one with Vedic *dadāti* and *dadhāti* in a minimal opposition. The words considered as whole strings, we can say there are two substrings which are the base of opposition between strings: *da-* and *-āti*. Moreover, there are two substrings, again exhibiting the minimality of opposition, equal to its elements, *d* and *dh*. Of course, with the segmentation of both words into morphs done correctly, there are two strings, each comprising three morphs: *da-dā-ti* and *da-dhā-ti*. The first and the last strings are in zero opposition, the middle strings differ in one element; in the case of this morph, the base is equal to *-ā-* and the complements are *d-* and *dh-*, respectively.

Generally, in the following lines we will mark any base of opposition as $\mathfrak{B}$, any complement as $\mathfrak{C}$. If necessary, complements and bases will be distinguished by numerical codes.

There are two possibilities for classification of oppositions between strings. One of them deals with a classification of relations between oppositions.

If the complement of the first string ($\sigma_1$) is equal to the complement of the second string ($\sigma_2$), i.e. $\mathfrak{C}_1 = \mathfrak{C}_2$ (and $\mathfrak{C}_2 = \mathfrak{C}_1$), both are in a **proportional** opposition. If the opposition between both (or more) strings is not proportional, such an opposition is **isolated** (see Trubetzkoy 1939: 63–66, Marcus 1967: 12–13, 80–82, Marcus 1969: 24–25, 79–80; compare Šefčík 2008a: 12). Using Hamming distance, in which equality is expressed by *0*s and inequality by *1*s, we can demonstrate this in the following table, proportional oppositions being on the left and isolated oppositions on the right:

| **proportional opposition** | | |
|---|---|---|
| $\sigma_1$ | $\sigma_2$ | H.-dist. |
| $\mathfrak{B}_1$ | $\mathfrak{B}_2$ | 0 or 1 |
| $\mathfrak{C}_1$ | $\mathfrak{C}_2$ | 0 |

| **isolated opposition** | | |
|---|---|---|
| $\sigma_1$ | $\sigma_2$ | H.-dist. |
| $\mathfrak{B}_1$ | $\mathfrak{B}_2$ | 0 or 1 |
| $\mathfrak{C}_1$ | $\mathfrak{C}_2$ | 1 |

**Note 7:** In following examples we respect the basic principles of the OIA morphology, however simplified.

**Example 13:** The words (= string of morphs) *gacchati* and *tiṣṭhati* have bases $\mathfrak{B}_1$ = *gaccha-* and $\mathfrak{B}_2$ = *tiṣṭha-* and complements $\mathfrak{C}_1 = \mathfrak{C}_2$ = *-ti*. So both are proportional.

**Example 14:** The words (= string of morphs) *gacchati* and *gacchasi* have bases $\mathfrak{B}_1$ = *gaccha-* = $\mathfrak{B}_2$ = *gaccha-* and complements $\mathfrak{C}_1$= *-ti* and $\mathfrak{C}_2$ = *-si*. Thus both are isolated.

If the base of the first string ($\sigma_1$) is equal to the base of the second string ($\sigma_2$), i.e. $\mathfrak{B}_1 = \mathfrak{B}_2$ (and $\mathfrak{B}_2 = \mathfrak{B}_1$), both are in a **homogeneous** opposition. If the opposition between both (or more) strings is not homogeneous, such an opposition is **heterogeneous** (see Trubetzkoy 1939: 60–63, Marcus 1967: 16–17, 80–82, Marcus 1969: 28–29, 79–80; compare Šefčík 2008a: 12). Using Hamming distances, in which equality is expressed by *0*s and inequality is expressed by *1*s, we can demonstrate this in the following table, homogeneous oppositions on the left and heterogenous oppositions on the right:

| **homogeneous opposition** | | |
|---|---|---|
| $\sigma_1$ | $\sigma_2$ | H.-dist. |
| $\mathfrak{B}_1$ | $\mathfrak{B}_2$ | 0 |
| $\mathfrak{C}_1$ | $\mathfrak{C}_2$ | 0 or 1 |

| **heterogeneous opposition** | | |
|---|---|---|
| $\sigma_1$ | $\sigma_2$ | H.-dist. |
| $\mathfrak{B}_1$ | $\mathfrak{B}_2$ | 1 |
| $\mathfrak{C}_1$ | $\mathfrak{C}_2$ | 0 or 1 |

**Note 8:** In following examples we respect the basic principles of the OIA morphology, however simplified.

**Example 15:** The words (= string of morphs) *gaccha-ti* and *gaccha-si* have bases $\mathfrak{B}_1$ = *gaccha-* = $\mathfrak{B}_2$ = *gaccha-* and complements $\mathfrak{C}_1$= *-ti* and $\mathfrak{C}_2$ = *-si*. So both word-strings are homogeneous.

**Example 16:** The words (= string of morphs) *gaccha-ti* and *tiṣṭha-ti* have bases $\mathfrak{B}_1$ = *gaccha-* and $\mathfrak{B}_2$ = *tiṣṭha-* and complements $\mathfrak{C}_1 = \mathfrak{C}_2$ = *-ti*. So both words are heterogeneous.

**Note 9:** Any opposition could be homogeneous and proportional only in case both complements and both bases are the same, i.e. $\mathfrak{C}_1 = \mathfrak{C}_2$ and $\mathfrak{B}_1 = \mathfrak{B}_2$ and hence $\sigma_1 = \sigma_2$. Such a opposition is known as an **opposition of equality**.

Another way of classifying oppositions is based on the relation between terms of an opposition.

If both bases of an opposition are equal (i.e. $\mathfrak{B}_1 = \mathfrak{B}_2$) and one complement $\mathfrak{C}_1$ is non-void and the second complement $\mathfrak{C}_2$ is void, such a opposition is **privative** (see Trubetzkoy 1939: 67, Marcus 1967: 4–7, 80–82, Marcus 1969: 17–22, 79–80, Brainerd 1971: 20–22; compare Šefčík 2008a: 12)

**Example 17:** The words (= strings of morphs) *deva* (= *deva-0*) and *deva-* are in a privative opposition.

If both bases of an opposition are equal (i.e. $\mathfrak{B}_1 = \mathfrak{B}_2$) and complements $\mathfrak{C}_1$ and $\mathfrak{C}_2$ are non-void (and hence not equal), such an opposition is **equipollent** (see Trubetzkoy 1939: 67, Marcus 1967: 7–9, 80–82, Marcus 1969: 20–21, 79–80, Brainerd 1971: 20–22; compare Šefčík 2008a: 13).

**Example 18:** The words (= strings of morphs) *deva-m* and *deva-* are in an equipollent opposition.

If both bases of an opposition are unequal (i.e. $\mathfrak{B}_1 \neq \mathfrak{B}_2$), or in other words, there is no common base of the opposition, and both complements $\mathfrak{C}_1$ and $\mathfrak{C}_2$ are non-void, such opposition is **disjunctive** (see Marcus 1967: 7–9, 80–82, Marcus 1969: 20–21, 79–80, Brainerd 1971: 20–22; compare Šefčík 2008a: 13).

**Example 19:** The words (= strings of morphs) *deva-m* and *kṣatriya-* are in a disjunctive opposition.

If both bases of an opposition are equal (i.e. $\mathfrak{B}_1 = \mathfrak{B}_2$) and complements $\mathfrak{C}_1$ and $\mathfrak{C}_2$ are void (i.e. there are no complements), such an opposition is **null** (see Marcus 1967: 4–7, 80–82, Marcus 1969: 17–22, 79–80, Brainerd 1971: 20–22; compare Šefčík 2008a: 12). This type of opposition is trivial but should be mentioned for the complete image.

**Example 20:** The words (= strings of morphs) *deva-m* and *deva-m* are in a null opposition.

## 4 Strings and gross metrics

Again, as oppositions between sets of values of phonemes could be metrized, using the fine metrics (see Šefčík 2008), oppositions between strings could be metrized, too, using the gross metrics.

We will slightly differ from pure formal application of metrics on strings, because in linguistics we must respect morphemic identity of given strings, i.e. partitioning of words on morphs (or morphemes, respectively). Pure formal application on abstract strings will be given in note below.

Considering identity as a zero distance in gross phonemic space and non-identity as at lest equal to minimal distance *1*, then we can use gross metrics on the strings.

Strings in the null opposition are in the mutual distance equal to *0*, because the mutual distance between bases is equal to *0* and mutual distance between components is equal to *0*, too.

| **null opposition** | | |
|---|---|---|
| $\sigma_1$ | $\sigma_2$ | |
| $\mathfrak{B}_1$ | $\mathfrak{B}_2$ | 0 |
| $\mathfrak{C}_1$ | $\mathfrak{C}_2$ | 0 |
| | | $\sum = 0$ |

Strings in the privative opposition are in mutual distance equal to *1*, because the mutual distance between both bases is null, but the distance between both complements is equal to *1*. It should be kept in mind that there is only one non-void complement against virtual (void) complement!

| **privative opposition** | | |
|---|---|---|
| $\sigma_1$ | $\sigma_2$ | |
| $\mathfrak{B}_1$ | $\mathfrak{B}_2$ | 0 |
| $\mathfrak{C}_1$ | $\mathfrak{C}_2$ | 1 |
| | | $\sum = 1$ |

Strings in the equipollent opposition are in the mutual distance equal to *2*, just for that reason that we interpret any equipollent opposition as equivalent to two privative oppositions. For reason of such qualification see (Šefčík – Osovský 2006, Šefčík 2008a, Šefčík 2008b).

| **equipollent opposition** | | |
|---|---|---|
| $\sigma_1$ | $\sigma_2$ | |
| $\mathfrak{B}_1$ | $\mathfrak{B}_2$ | 0 |
| $\mathfrak{C}_{1a}$ | $\mathfrak{C}_{2a}$ | 1 |
| $\mathfrak{C}_{1b}$ | $\mathfrak{C}_{2b}$ | 1 |
| | | $\sum = 2$ |

Strings in the disjunctive opposition are in the mutual distance equal or greatest than *3*. The reasons are obvious and analogous to the previous example of equipollent opposition. Both (or more) complements are in the same mutual relation as complements of equipollent opposition, besides there always is difference in bases, because no common base means minimal distance of value equal to *1*.

| **disjunctive opposition** | | |
|---|---|---|
| $\sigma_1$ | $\sigma_2$ | |
| $\mathfrak{B}_1$ | $\mathfrak{B}_2$ | 1 |
| $\mathfrak{C}_{1a}$ | $\mathfrak{C}_{2a}$ | 1 |
| $\mathfrak{C}_{1b}$ | $\mathfrak{C}_{2b}$ | 1 |
| | | $\sum = 3$ |

**Note 10:** An application of gross metrics on strings without any regard to linguistic reasons for classification of strings on substrings, i.e. limitation of words on morphs as least possible unions of form and meaning as given above, leads toward more simply classification. For complexity of our sketch we will point out some important consequences in application.

First, we should keep in mind, that in such case it is not possible to do distinction between privative and equipollent oppositions, because first is only in contrast of non-void complement to a void complement, second is between two non-void complements. Hence the distance between strings both in privative or equipollent oppositions is equal to *1*, because there is only one difference (in complements). We can mark both types of opposition as **inclusive** opposition (inclusive for common base of opposition). Subsequently, the disjunctive (for difference here termed **exclusive**) opposition is hence equal to *2* for same reasons; there is one difference in the bases and another in complements:

| **inclusive opposition** | | |
|---|---|---|
| $\sigma_1$ | $\sigma_2$ | |
| $\mathfrak{B}_1$ | $\mathfrak{B}_2$ | 0 |
| $\mathfrak{C}_1$ | $\mathfrak{C}_2$ | 1 |
| | | $\sum = 1$ |

| **exclusive opposition** | | |
|---|---|---|
| $\sigma_1$ | $\sigma_2$ | |
| $\mathfrak{B}_1$ | $\mathfrak{B}_2$ | 1 |
| $\mathfrak{C}_1$ | $\mathfrak{C}_2$ | 1 |
| | | $\sum = 2$ |

## 5 Final statements

In lines above we tried to show, how oppositions in the sense of description by Trubetzkoy and Marcus could be successfully applied not only on phonemes, but on whole strings of phonemes, from which most important are those equal to basic morphemic elements, i.e. morphs and morphemes.

Any opposition could be metrized and the metrics, kind of metric space used above is termed gross metrics.

## Literature

BRAINERD, B. (1971): *Introduction to the Mathematics of Language Study*. New York: American Elsevier

CANTINEAU, J. (1952): Les oppositions significatives. *Cahiers Ferdinand de Saussure* 10, 11–40

CANTINEAU, J. (1955): Le classement logique des oppositions. *Word* 11/1, 1–9

HAMMING, R. W. (1950), Error Detecting and Error Correcting Codes, *Bell System Technical Journal* 26(2), 147–160

MACDONELL, A. A. (1910): *Vedic Grammar*. Strassburg: Karl J. Trübner

MACDONELL, A. A. (1916): *A Vedic Grammar for Students*. Oxford: Clarendon Press

MARCUS, S. (1967): *Introduction mathématique à la linguistique structurale*. Paris: Dunod

MARCUS, S. (1969): *Algebraické modely v lingvistice*. Praha: Academia

ŠEFČÍK, O. – OSOVSKÝ, M. (2006): Zobrazení mezi fonologickými komponenty. *SPFFBU A* 54. 19–29

ŠEFČÍK, O. (2007): K užití metriky ve fonologii. *SPFFBU A* 55. 19–26

Šefčík, O. (2008a): Values, features, fine metrics and oppositions. *SPFFBU A* 56. 6–14
Šefčík, O. (2008b): On significance of alternations for functioning of a phonological system. *Slavia* 77/1. 171–176
Trubetzkoy, N. S. (1939): *Grundzüge der Phonologie.* Prague: TCLP

## ŘETĚZCE, OPOZICE, METRIKA

V tomto příspěvku se věnujeme některým otázkám, spojeným s pojmem hrubé metriky, tj. takové metriky, jakou použijeme na vyjádření vzdálenosti mezi řetězci fonémů a s tím spojenými otázkami opozic mezi řetězci vůbec. Článek souvisí s předchozími články zabývajícími se teorií opozic a metrických prostorů ve fonologii.

V článku vycházíme z pojmů a postupů poprvé zavedených Marcusem, zavádíme definici řetězce fonémů, hrubé metriky vytvořené na takových řetězcích a aplikujeme opozice ve smyslu definic Trubeckého, Cantineaua, Marcuse a Brainerda.

Příklady jsou převzaty z védského sanskrtu.

*Ondřej Šefčík*
*Ústav jazykovědy a baltistiky*
*Filosofická fakulta, Masarykova universita*
*A. Nováka 1*
*602 00 Brno*
*e-mail: sefcik@phil.muni.cz*