

3.

Co-reference¹

ANDREA NAGY—FRANCISKA SKUTTA

1. Co-reference means the relation ensuring the textual connectivity of two or more linguistic elements with different verbal manifestations which, according to the interpreter, refer to the same entity of the given text world. As to the analysis of the relation between the elements which are considered to be coreferential, i.e. co-reference analysis, JÁNOS S. PETŐFI and EDIT DOBI (1998: 238) write the following:

“The *analysis of co-reference* must take into consideration the relation of two levels: (a) one is the interpreter’s mental image of the ‘world fragment’ represented in the analysed text, with its entities (i.e., with the mental images of persons, objects, concepts, attributes, events, etc.); (b) the other is the physical manifestation of the analysed text, with the text components interpreted syntactically / semantically by the interpreter. A part of these text components refers to entities of the supposed ‘world fragment’; the other part expresses diverse statements about the entities referred to.

A *co-reference relation* (‘together-referring relation’) is nothing else than a sort of ‘affinity’ relation between the text components supposed to refer to the same entity of the presumed world fragment. This relation can be ‘identity’ (such a relation exists between — either identical or different — text components which presumably refer to the same entity of the presumed world fragment); it can be in a certain sense that of ‘possessor-possession’ (this holds, among others, for two text components, one of which refers to a person or an object, while the other refers to an entity which can be considered as one belonging to this person or object), but it can be of several other types as well.”

2. A particularly important question in the case of co-reference analysis from a text linguistic / textological point of view is the most effective way of representing explicitly the coreferential elements and their mutual relations in a given text, as well as the deducible information based on linguistic and general knowledge about the world, both of which are necessary for revealing the coreferential relations in the analysed text.

¹ This publication was supported by the TÁMOP-4.2.3-08/1-2009-0017 project. The project was co-financed by the European Union and the European Social Fund.

“A polyglot research program in textology / text linguistics” has provided for these question-raising theoretical and practical problems a possible solution, namely the co-reference analysis by JÁNOS S. PETŐFI (1997: 24-38), (1998: 15-31), a highly inspiring model for co-reference analyses of texts belonging to different types.

In fact, this basic method of analysis has made it possible to compare directly the results of the co-reference analysis of texts belonging to different types of texts.

In our essay, we shall first present the model of JÁNOS S. PETŐFI (1998: 15-31); our presentation is based on his study published in *Officina Textologica 2* and will be followed by that of related articles on co-reference.

The sample text analysed in the orientating article is a biblical text extracted from *New Testament*:

Ve: Mt. 9, 9-13. The Calling of Matthew²

⁹As Jesus passed on from there, he saw a man called Matthew sitting at the tax booth. He said to him, “Follow me.” And he rose and followed him.

¹⁰And as Jesus reclined at table in the house, behold, many tax collectors and sinners came and were reclining with Jesus and his disciples. ¹¹And when the Pharisees saw this, they said to his disciples, “Why does your teacher eat with tax collectors and sinners?” ¹²But when he heard it, he said, “Those who are well have no need of a physician, but those who are sick.” ¹³Go and learn what this means, “I desire mercy, and not sacrifice.” For I came not to call the righteous, but sinners.

János S. Petőfi resumes the expedient steps of co-reference analysis as follows:

“*First*, I assign — giving up the biblical verse numbering — codes in square brackets (see for example [C01]) to the text sentences to identify them. The letter ‘C’ of these codes stands for the expression ‘first-degree macro-compositional unit’, while the number indicates that given to a text sentence in the order of sentences in the text. A text sentence is defined as a unit beginning with a capital letter and closed with a full stop, a colon, an exclamation mark or a question mark.

As the *second* step, I complete each text sentence with verbal expressions (in italics) which refer to elements / relations deducible from the sentences themselves, from their verbal context and / or from our general knowledge about

² English Standard Version, <http://www.biblegateway.com>. – The original analysis uses the Hungarian translation of the *Bible*.

the world in a way that each text sentence shall constitute an independent ‘information unit’, while coreferential relations inside and between the text sentences must be expressed explicitly. In order to indicate that in the analysed text sentence I consider an expression consisting of more than one elements as a single component, I use the concatenation symbol “^”. I assign to the text sentences completed with verbal expressions a code completed with the symbol “/&vb” (see for example [C01/&vb]).

As the *third* step — under the title “Comment”— I analyse both the inner syntactic structure and the coreferential relations of each text sentence completed with verbal elements. (However, for the sake of simplicity, I do not deal with articles, negative particles, time adverbials, tenses, and conjunctions.) [...]

As the *fourth* step, I determine which elements of the text sentence completed with verbal expressions should be represented by co-reference indices, I create these indices adding a comment to them; for indexation, I use a code consisting of the letter “i” and of a two-digit number (see e.g. **i03** or **i11**).

As the *fifth* (and last) step, I finally create for each text sentence its variant completed with co-reference indices. After their original code, I assign to the text sentences completed with co-reference indices the symbol “/&ind” (see e. g. [C01/&ind]). I create the variants with co-reference indices based on the analysis of the text sentences completed with verbal expressions in the following way:

- I place co-reference indices each time in square brackets after the expression / word / suffix to be marked by them,
- I put co-reference indices in bold type when they represent nominal expressions at their first occurrence and I put an equal mark before them,
- I use normal characters for co-reference indices in all other cases, without an equal mark.” (PETŐFI 1998: 15-17)

3. We shall now illustrate the analysis described above. First, we determine the text sentences of the sample text and assign a code to each of them.

Ve[C00]–[C12]:

[C00]*Mt. 9, 9-13. The Calling of Matthew*

[C01]As Jesus passed on from there, he saw a man called Matthew sitting at the tax booth. [C02]He said to him, [C03]“Follow me”. [C04]And he rose and followed him. [C05]And as Jesus reclined at table in the house, behold, many tax collectors and sinners came and were reclining with Jesus and his disciples. [C06]And when the Pharisees saw this, they said to his disciples, [C07]“Why does your teacher eat with tax collectors and sinners?” [C08]But when he heard it, he said, [C09]“Those who are well have no need of a physician, but those who are sick. [C10]Go and learn what this means, [C11]‘I desire mercy, and not sacrifice.’ [C12]For I came not to call the righteous, but sinners.

As the next step, we have to complete the text sentences with the information deducible from the verbal context and from our general knowledge about the world. Let us take text sentences [C01]–[C03] as an example.

[C01] As Jesus passed on from there, he saw a man called Matthew sitting at the tax booth.

[C01/&vb]:

As Jesus passed on from there (*from the place where Jesus previously was*), he (*Jesus*) saw a[^]man[^]called[^]Matthew (*who was*) sitting at the[^]tax[^]booth.

[C02] He said to him,

[C02/&vb]:

He (*Jesus*) said (*the following*) to him (*to the man called Matthew*),

[C03] “Follow me”.

[C03/&vb]:

(*You, the man called Matthew*) Follow me (*Jesus*).

After the detailed analysis under the title “Comment”, the following co-reference indices can be assigned to the elements of the text sentences made explicit by verbal expressions:

i01 = Jesus,

(*i02 = the place where Jesus previously was*),

i03 = the man called Matthew,

i04 = the tax booth.

As the final step of the analysis, we assign the above co-reference indices to the expressions of the text sentences, thus replacing the explicit verbal information.

Ve/&ind: *Mt. 9, 9-13. The Calling of Matthew* [C01]–[C03]

[C01/&ind]:

As Jesus[=**i01**] passed[i01] on from there(**i02**), he[i01] saw[i01] a[^]man[^]called[^]Matthew[=**i03**] sitting[i03] at the[^]tax[^]booth[=**i04**].

[C02/&ind]:

He[i01] said[i01][**C03**] to him[i03],

[C03/&ind]:

Follow[i03] me[i01].

The analytical steps continued in this way, i.e., by examining text sentences one by one, result in the complete list of co-reference indices in a given text. The following list shows the co-reference indices of the text analysed here:

i01 = Jesus,

(*i02 = the place where Jesus previously was*),

i03 = the man called Matthew,

i04 = the tax booth (the place at the tax booth),

i05 = the house of the man called Matthew = the house(i03)[=**i05**],

i06 = the many tax collectors (coming to the house of Matthew),

- °i07 = the many sinners (coming to the house of Matthew),
- *i08 = the disciples(i01)[=i08] of Jesus (in the house of Matthew),
- i09 = table/tables (in the house of Matthew),
- i10 = the [’(the many tax collectors and the (many) sinners) with Jesus and his (Jesus’) disciples together (in the house of the man called Matthew) reclining at table)’ event-seeing] Pharisees,
- i11 = those who are well (in general),
- i12 = those who are sick (in general),
- i13 = a physician (in general),
- i14 = (unspecified) mercy,
- i15 = (unspecified) sacrifice,
- i16 = the (unspecified) righteous.

The distribution of coreferential elements in the text sentences can also be represented in the form of a table clearly indicating the co-reference relations in the given text. The co-reference relations of text sentences [C01]–[C03] analysed above are represented by the following table:

	[C01]	[C02]	[C03]
i01	Jesus, passed on, he, saw	he, said	me
<i>i02</i>	<i>(from the place)</i>		
i03	the man called Matthew, sitting	to him	Follow
i04	at the tax booth		



It is then along the main lines of this orientating article that further investigations in co-reference have been carried out by the contributors of “A polyglot research program in textology / text linguistics”. Relevant studies are to be found in several issues of *Officina Textologica*, the most important volume in this respect being precisely the one — *Volume 2* (1998) — containing Petőfi’s leading article and eleven other studies on aspects of co-reference. In fact, under the title *Coreferential elements, co-reference relations*, the volume offers analyses of a wide range of phenomena concerning the realization of co-reference in different text types represented either by complete short texts or by extracts from longer ones. This first thematic collection of articles was then followed by a so-called “discussion” volume — *Volume 4* (2000) —, which

raises some more theoretical issues related to co-reference. While these two volumes analyse coreferential elements and relations exclusively in Hungarian texts (or, occasionally, in texts translated into Hungarian), the six essays of *Volume 12* (2005) examine problems of co-reference in a contrastive approach, comparing Hungarian with English, French and German, respectively. Finally, several other issues of *Officina Textologica*, though not necessarily focussing on co-reference, contain articles which occasionally touch upon questions of connectivity and cohesion. Since it would be impossible to give a detailed presentation of each study in the framework of this short essay, and in order to avoid repetition, we shall discuss certain *problems* of co-reference analysis rather than present individual articles. Owing to the great variety of coreferential phenomena and of analytic methods, we shall concentrate on *Volume 2* and examine how different text types may influence, or even command, the linguistic formulation of co-reference relations.

As we know, texts can be classified into types, which are best defined as a set of features selected according to different criteria, both linguistic and pragmatic. In fact, from a typological point of view, the texts described in the eleven articles that follow the Petőfi model in *Volume 2* can be divided in the first step into non-literary and literary texts, each of these two general types being represented here by specific genres.

Non-literary texts — conceived for usage in various every-day situations and interactions — are exemplified in this volume by certain genres which are embodied either in texts with usually reduced dimensions, or on the contrary, in texts of considerable length like scientific or scholarly books. To illustrate the former group, the authors quote and analyse such complete short texts as an *encyclopaedia entry* (3, DOBI), a *recipe* (4, DOMONKOSI), a *business letter* for advertising purposes (5, CS. JÓNÁS), and a *newspaper article* (11, SZIKSZAI NAGY), whereas the latter type appears in one article on a *scholarly text*, represented by an extract from a book on literary theory (6, SKUTTA). A special case of non-literary texts, namely a new type of *multimedia product*, the CD-ROM with its *Help* feature, is also dealt with (8, BENKES—VASS).³

Literary texts proper appear in three articles, one analysing a *poem* (10, B. FEJES), another a short extract from a *novel* (9, BÉKÉSI), and the third an “entry” from a so-called “*dictionary-novel*” (12, TOLCSVAI NAGY). Finally, two borderline cases of literature are treated in two studies respectively: a *folk-tale* (7, TUBA) and a biblical text, an extract from *New Testament* (2, BODA—PORKOLÁB).

³ The number in brackets (preceding here the name of the author and later used without a name) indicates the number in the order of articles in *Volume 2*. See the list of articles below.

As it may be seen from the above list, the corpus used for the analysis of co-reference is sufficiently varied for possibly representing a considerable amount of phenomena which can be examined from the point of view of textology / text linguistics. It seems that a comprehensive survey of those phenomena — as the one by PETŐFI—DOBI at the end of *Officina Textologica 2* (pp. 238-261) — has to answer the following questions:

(1) What kind of *world entities* can be referred to by means of specific referential units of the text?

(2) What sort of elements can be used as *referential text units*?

(3) What kind of *co-reference relations* can be identified, in other words, what sort of referential differences may be included in the notion of co-reference?

(4) Which is the most effective way of *representing* coreferential elements and co-reference relations?

In what follows, we shall discuss these four questions one by one, with the answers suggested by the presentation of PETŐFI—DOBI and the analyses in *Volume 2*. It is important to bear in mind, however, that an exhaustive description of all relevant phenomena would need an investigation on a much larger corpus.

(1) *World entities* referred to by linguistic means are either countable or uncountable and both classes contain further subdivisions.

Countable world entities are persons, objects, etc.:

(a) numerically well-defined and clearly identifiable:

Jesus (1), Emperor Nero (3), Thomas Mann (10), Levin (9), Vronsky (9)

the king (7), the party leader (11), the mother (9), the coachmen (3)

the chicken (4), the cock (7)

the house of the man called Matthew (1), the red VW Golf (5), the gold coin (7)

the large and high mountain where the angel led John (2), the Holy City (2)

the Roman Empire (3), the *Circus Maximus* (3)

the Renaissance (6)

the two aspects (6), the three main parts (8)

(b) numerically well-defined but not clearly identifiable:

a royal family (7), the 150 to 200 thousand people watching the race (3)

an unspecified CD-ROM (8)

(c) numerically indefinite but clearly identifiable:

the tax collectors and the sinners in the house of Matthew (1), the servants (7)

tables in the house of Matthew (1), the products of our firm (5)

the written records on chariots races (3), the works of the classical authors (6)

(d) numerically indefinite and not clearly identifiable:

people (7), those who are well (1), those who are sick (1)

the loveliest fruit trees (5), the chariots (3)

unspecified words (8), the conditions of the genesis of (literary) works (6)

Uncountable entities are:

(a) of a definite quantity and clearly identifiable:

the length / width of the *Circus Maximus* (3)

(b) of a definite quantity but not clearly identifiable:

150 g of mushrooms (4), 1 spoonful of parsley (4)

(c) of an indefinite quantity but clearly identifiable:

the material of the walls of the Holy City (2), the water in the well (7)

(d) of an indefinite quantity and not clearly identifiable:

glass (2), a pinch of salt (4), water (4)

Entities *different* from the above are of six main types:

(a) objects whose consistency is continually changing:

1 chicken (4), 150 g of rice (4)

(b) abstract notions:

mercy (1), sacrifice (1), literature (6), the idea of duality (6)

(c) linguistic / metalinguistic entities (titles, names, forms of address):

the encyclopaedia entry "Chariots race" (3), Dear Mrs. Szegedi (5), Princess (7)

(d) qualities:

of childhood (12)

(e) states of affairs:

(sb) asks (sb to do sth) (10)

(f) time, period:

again (11), the time of the tale (7)

(2) Types of *referential text units*

(a) proper names:

Jesus (1), Hans Castorp (10), the Roman Empire (3)

(b) nouns, noun phrases:

the bride (2), the little coin (7), the history of literary research (6), in the dark (10)

(c) pronouns (personal, demonstrative, relative, etc.):

he (12), me (1), with us (10), these (8), in that (8), who (12), which (2)

(d) [in Hungarian] personal possessive suffixes (referring to the possessor):

my name (12), our catalogue (5), the beginning of [it =] the race (3)

(e) adverbs:

thus (2), then (7), again (11)

(f) [in Hungarian] verbal prefixes:

you give it back (7), it flew up (7)

(g) [in Hungarian] verbal suffixes (referring to the grammatical person):

I know (12), he saw (1), you can choose (5), take a seat (10)

(h) infinitive:

to resemble (9), to remember (11)

- (i) finite forms of verbs as autonomous states of affairs:
(sb) asks (sb to do sth) (10), (sb never) lies
- (j) non-linguistic symbols:
my ~ [for *ideal*, an entry in the so-called dictionary-novel] (12)
- (3) Types and limits of *co-reference relations*
 - (a) complete referential identity (repetition, synonymy, pronominal substitution, indication of the person of the possessor, conjugation, appositive — occasionally defining — constructions):
Jesus, he, me, him, his disciples, your teacher, I (1)
 - (b) hyponym — hyperonym relations:
hyperonym: all sorts of precious stones — hyponyms: jasper, emerald, topaz (2)
 - (c) associative relations:
his photo — light (11)
 - (d) set — subset relations:
the 150 to 200 thousand people watching the race — those making a bet (3)
 - (e) change of state (of consistency) of the entity referred to:
the chicken, the chicken cut to pieces, stew it, spice it, serve it (4)
 - (f) metalinguistic correspondence:
chariots race [encyclopaedia entry] — speed contest (3)
 - (g) reference to text segments:
tax collectors and sinners [...] reclining with Jesus and his disciples — this (1)
 - (h) co-reference between a noun phrase and its central semantic constituent:
the literary work(s) — literature (6)
 - (i) co-reference ensured by the given communicative channel:
tell us a tale — your words — you speak — we listen (10)

(4) Ways of *representing* coreferential elements and co-reference relations

At the present state of research, it is impossible to give a definitive answer to the question concerning the most effective way of representing coreferential elements and co-reference relations. While based on the representational system used in the orientating article, the analyses of different text types suggest various additional solutions specific to those types of texts. Since the use of indices for representing simple noun phrases (e.g. proper names or a noun preceded by a determinant and interpreted in its concrete meaning) does not seem to cause any difficulty, we shall focus below on certain problematic cases of indexation. In fact, some of those phenomena are specific to the Hungarian language, such as the existence of two conjugations — “subjective” and “objective” —, as well as of verbal prefixes and of personal possessive suffixes (added to nouns), while a considerable number of phenomena are clearly translinguistic, among others forms of address, or reference made either to a complex noun phrase or to one of its constituents only. In what follows, we shall briefly comment on a series of “problematic issues”, raised by the analyses in *Officina Textologica 2*.

I. *Types of referential text units*– *Personal suffixes in the objective conjugation*

These suffixes ensure a double co-reference, i.e., with the subject and with the definite direct object of the verb. Both co-reference indices — the first representing the subject, the second the object, and being separated by the sign — should be marked immediately after the conjugated form of the verb:

The cock [i01] gave [i01 i02] the coin [i02] [to the Princess] (7).

– *Verbal prefixes*

Since Hungarian verbal prefixes can eventually refer to the participants of the action designated by the verb, in such cases they should be supplied with co-reference indices of their own. The following English translation cannot reproduce exactly the original verbal prefixes, which are then represented by the preposition *to* and the adverbial particle *back*:

I [i01]’ll give [i01] it [i02] to you [i05], Princess [i05], but only if you [i05] give [i05] it [i02] back [i01] (7).

– *Personal possessive suffixes*

In Hungarian, nouns can take so-called “personal possessive suffixes”, which refer to the possessor and, at the same time, indicate the grammatical person of the possessor referred to in a particular communicative situation. Thus, functionally, they correspond to English possessive determinants (*my, your, his*, etc.), but since in Hungarian the possessive reference is incorporated in the noun, the latter will be followed by a complex co-reference index, the first member of which represents the referent designated by the noun, and the second refers to the possessor expressed by the suffix:

the king [i04]’s sister [i05 +i04], an elderly princess [i05] (7).

– *Title, name, social rank, forms of address*

Such phrases can be used either to refer to world entities or to quote text segments, but when a co-reference chain occurs in the text, the *title / name / social rank / forms of address* phrases normally refer to world entities:

chariots race [entry] — speed contest [world entity] (3)

She [world entity] was called *Princess* [name], because her [world entity] brother [world entity] was a *king* [rank] (7).

The Princess saw it [the cock = world entity] and said: ‘My dear *cock*’ [form of address] (7).

Such phrases, when used as “quotations”, should be marked with special additional symbols (, * , ’).

– *Concrete and abstract meanings of a word*

A sort of “type / token” relation holds for words with both concrete (token) and abstract (type) meanings, such as:

a [particular] gold *coin* found by the cock — *coin* [‘money’ in general] (7).

Such phrases, when used in their abstract meaning, should be marked with special additional symbols (*).

– *Reference made to a text sentence or to a larger text segment*

This type of reference is often realized by means of a demonstrative pronoun (*this, that*); indexation can resume the whole sentence / segment referred to by use of a code standing for the macro-compositional unit in question:

[C05]And as Jesus reclined at table in the house, behold, many tax collectors and sinners came and were reclining with Jesus and his disciples.

[C06]And when the Pharisees saw *this*[C05], they said[C07] to his disciples,

[C07] “Why does your teacher eat with tax collectors and sinners?” (1).

II. *Types and limits of co-reference relations*

– *Change of state of the entity referred to*

Each phase of a change of state can be represented in a complex way, by means of the constant index standing for the entity and the codes (in subscript) of text sentences describing the actual states of the entity:

[C02] *the chicken* [i02] cut to pieces — *the roast chicken* [i02C02] (4).

– *Set — subset relations*

In order to show that the validity of a predicative statement does not cover the whole set, but extends only to a subset of the whole, an additional symbol may be introduced (“ \subset ” = ‘subset’):

the 150 to 200 thousand people [i07] watching the race — *those* [\subset i07] making a bet (3).

– *Associative relations*

It is possible to conceive different representations of associative relations — as those between *photo* and *light* (11)⁴ —, but it is advisable to use one and the same index for the common semantic feature, accompanied by additional symbols, e.g. subscripts.

– *A noun phrase and its central semantic constituent*

In order to show the presence of the same semantic constituent in two formally different noun phrases, it is possible to use the same numerical index accompanied by two different letters:

the literary [j01]^work(s) [i01] — literature [j01] (6).

– *Co-reference indices whose interpretation implies other referential elements*

In such cases an additional symbol (e.g. *) can indicate the presence of another — autonomous — referential element in the text, which is necessary for the interpretation of the referential unit in question:

⁴ In Hungarian the word for *photo* is a compound containing the word for *light* followed by the one for *picture*: “light-picture”. The associative relation is thus quite explicit. – Study (11) uses an indexation different from that of the orientating article, which would be too long to explain in a short comment.

the *Circus Maximus* [i19] — (length [i19][i20*], width [i19][i21*]) (3).

III. *Simple referential (not coreferential) index*

It is advisable to use a special symbol (e.g. **) for indicating the single occurrence of a referential element which does not enter a co-reference chain:

antiquity [i04**] (3).



Having examined various problems in connection with (1) world entities referred to, (2) referential text units, (3) types and limits of co-reference relations, and (4) the most effective ways of representing coreferential elements and co-reference relations, we shall now turn briefly to the twelve articles of *Officina Textologica* 2, in order to treat some special questions raised by the types of texts analysed in this volume.

By giving an explicit co-reference analysis of a short extract from *New Testament*, the orientating article (1) by JÁNOS S. PETŐFI offers a model applicable for any text type, with minor changes in the representation of coreferential elements and co-reference relations. That is certainly due, among others, to the nature of the analysed text itself: in fact, *The Calling of Matthew*, Mt. 9, 9-13 is a sort of “unmarked” text, an almost everyday *narrative* sequence with a relatively simple, chronological time structure and with a small number of actors — Jesus, Matthew, the disciples, the Pharisees, the tax collectors and the sinners coming to Matthew’s house —, whose multiple textual occurrences (in the form of noun phrases and pronouns) create well-defined co-reference chains. Thus the explicit representation of those chains by means of coreferential indices remains unequivocal throughout the passage (for details, see above).

The authors of Study (2), KÁROLY I. BODA and JUDIT PORKOLÁB, examine another extract from *New Testament*, namely a passage presenting the New Jerusalem in the *Revelation to John* (*Rev. 21, 9-23*). Though set in a narrative framework — one of the seven angels coming to John, carrying him away to a high mountain and showing him the Holy City — John’s vision is rendered in a *descriptive* sequence, with a spatial organization of the City and its parts, the wall, the twelve gates and the street. Whereas in the *Calling of Matthew* co-reference chains represent first of all human beings performing actions in time, indices identifying the elements of John’s vision in the *Revelation* stand for static objects characterized by measures, forms, materials and colours. Since the vivid description of the radiant city implies emphatic repetition of the names of its parts, the extract lends itself to an analysis slightly different from that of the orientating article: the authors have invented a frequency matrix for representing the textual occurrences of referential units.

Study (3) by EDIT DOBI deals with a special type of short text, namely an *encyclopaedia entry*, ‘chariots race’, extracted from the *Hungarian Larousse Encyclopaedia*. This entry combines characteristics of narrative and descriptive texts, for it *describes* a social *event*: in fact, the chariots race, though occurring in time, will not turn into a *story*, and thus the presentation has to focus on its general features and circumstances. Given the concise *informative* and *definitive* nature of encyclopaedia entries, the analysis of co-reference relations is concerned with several specific problems, among others the relationship between the referential and the metalinguistic interpretations of a noun phrase and, in the latter case, the use of a special symbol; divisions of the entry and definitions; abbreviations and non-linguistic symbols. Questions of a more general validity concern the identification and the separation of text sentences, especially of those that form the typical introduction to the entry. Following the definitive nature of this type of text, the concept of “partial co-reference” and the possibility of its representation are also examined.

In some respects, Study (4) by ÁGNES DOMONKOSI raises issues similar to those relevant for the textual organization and the visual manifestation of an encyclopaedia entry. Here the selected text, a *recipe*, ‘chicken with rice’ (from Lukács Túrós: *Cookery Book for Maidens and Women*), is also a short text (of about the same length as the entry) with a title and a typographical distinction of text segments, so that these aspects of the text must be taken into consideration in the course of referential indexation (e.g. metalinguistic interpretation, the coreferential status of ‘ingredients’). Furthermore, the recipe also describes a process occurring in time, by giving instructions for a series of operations to be performed for the realization of a goal. However, as opposed to the entry — and in fact, to all other texts examined in this volume —, the recipe refers to world entities continuously changing in the process of cooking, which raises the problem of ‘identity’ / ‘difference’ and, consequently, of coreferential indexation. The solution suggested here marks each phase of the alteration of the entity prior to its state under examination.

With Study (5) by ERZSÉBET CS. JÓNÁS, we turn to a type of text rather different from the previous types in one respect, namely as far as reference to the text recipient is concerned. In fact, the first three texts analysed in the volume make no explicit reference to the reader. As for the recipe, its potential readers / performers are at least implied in the text by means of verb forms (1st person plural). On the contrary, the text analysed here, a *business letter* written for advertising purposes (i.e., each copy of it) is clearly addressed to a particular recipient, using names, typical formulas of politeness and often the imperative of the verbs, with which the sender tries to elicit a positive response to the offer. The analysis therefore takes into account the sociological and communicative aspects of the business letter, as well as the convincing power of the text. The

description of co-reference relations faces some text specific problems, such as the different representations of and the constant feedback to the addressee.

As compared to the previously analysed texts, in which, for the most part, new arguments introduce new world entities, the text examined in Study (6) by FRANCISKA SKUTTA represents a type which aims at describing, in *specialist language*, different aspects of one and the same object, often an abstract domain. The author analyses a short extract — the first two paragraphs — from a comprehensive book on literary theory (Elemér Hankiss: *Literary Work as Complex Model*), and wishes to show some genre-specific characteristics of scholarly writing, which have an impact on co-reference relations. It occurs that in *argumentative* texts referential units forming a co-reference chain are often linked by partial co-reference of different types (set / subset, general / particular, abstract / concrete). Furthermore, the text abounds in complex noun phrases of which only one element enters a co-reference chain, a problem that needs special attention: it is namely advisable to elaborate a way of representing the central and peripheral elements by two distinct, though related, indices.

Study (7) by MÁRTA TUBA analyses an extract from a narrative text rather different in style from the previously quoted texts: it is namely a Palóc *folk-tale* (from the North of Hungary), *The Cock's Gold Coin*, told in a somewhat crude popular language. The syntactic structures of text sentences being relatively simple, with a certain number of repetitions characteristic of folk-tales, the author focuses on some other text-specific features. Since the strongest organizing factor of this type of narrative is undoubtedly the flow of time, the chronological succession of events, all adverbials of time are represented by one common index: “i00”, which is a new element added to the analysis suggested in the orientating article, showing the key role of time in the tale. The representation also gives an account of the frequent double functioning of the same expression (usually a noun phrase), used either as a form of address, or as the indication of a social rank — in both cases, the same index appears with a supplementary symbol to distinguish between the two interpretations.

Study (8) is an exception among the articles of this volume in that it deals with a type of communication that is not (exclusively) linguistic. The authors, RÉKA BENKES and LÁSZLÓ VASS, examine, within the framework of semiotic textology, a multimedia product, *Hungarian Bookshelf* CD-ROM, and in particular, its *Help* feature. This medium is composed of verbal and non-verbal — iconic — elements, the totality of which can be considered as a *hypertext*, which has a non-linear organization, but at the same time needs a linear manifestation when it is interpreted as “instructions for use”. The analysis of the linear manifestation of the hypertext requires certain interpolations for an explicit description of co-reference relations established between verbal and

iconic text components, the latter including non-linguistic symbols, colours and typography.

Study (9) by IMRE BÉKÉSI analyses a short extract from L. Tolstoy's *Anna Karenina* (First part, XII, 3-4; Hungarian translation by László Németh). Although the article deals with a *novel*, what interests the author is not so much the narrative aspect of the text, but rather the *argumentative* role of the so-called *key sentence*, which resumes and prefigures the semantic content of a paragraph (or a larger text segment, eventually the whole novel), while the latter gives a detailed explanation of the situation having been referred to cataphorically. Coreferential indexation shows explicitly that the key sentence contains lexical reference to all essential world entities that form co-reference chains in the following text segment. In order to demonstrate the argumentative character of the key sentence and of its relationship with the following text segment, the author also analyses the logical macro-structure of the whole extract organized according to relations of explication-deduction and opposition.

Study (10) also deals with a *literary* text: KATALIN B. FEJES examines a poem, *The Greeting of Thomas Mann* by Attila József, and insists on a special characteristic of this text, which distinguishes it from all the other texts presented in the volume. As it is indicated in its title, the poem depicts an imaginary *communicative situation*, in which the poet speaks on behalf of the audience, “we”, and addresses the guest, “you”, by a series of verbs in the imperative form, performing the global speech act of “request”. The imagined situation of oral communication implies the simultaneous functioning of three channels, auditory, visual and kinetic, each of which is considered here as an autonomous level of co-reference chains. The specificity of coreferential representation in this article is that not only world entities, but even particular acts of request are designated by coreferential indices, since it is those acts, inspired by the situation and the poet's emotions, that form the core of the structure of the poem.

Study (11) by IRMA SZIKSZAI NAGY presents a short *newspaper article*: ‘Bike-party leader’ (*Kurír*, VII [1997], 5, p. 16). The Hungarian title of the article contains a pun on the word “party” (used here in the political sense), and suggests an ironic comment on a small act of charity by a well-known politician. Taking into account the special effect of the tone of the text, the author combines coreferential and stylistic analyses, thus offering a description partly different from that proposed in the orientating article. Apart from a few slight modifications (e.g., a simplified notation) in the representation of co-reference relations — Arabic and Roman numerals for text sentences and for coreferential elements respectively —, some more profound changes are introduced in the analysis, in order to show, in several diagrams, the various textual manifestations of co-reference chains built on dominant lexical units. The author

also studies the distribution and the relative frequency of coreferential elements, both nominal and verbal, forming “clusters”, and having an impact on the style of the text.

The last analysis of this volume, Study (12) by GÁBOR TOLCSVAI NAGY, is again concerned with a *literary* text, but a very special one, a so-called “dictionary novel” by Ferenc Temesi, *Dust*, article: ‘ideal’. Thus, this passage partly bears the characteristics of an *encyclopaedia entry*, as the one presented in Study (3), except for the personal involvement of its narrator and the slightly mocking tone he uses to speak about his childhood ideal and the disillusion of adulthood. Some of the coreferential phenomena are similar in the two entries, for instance the use of non-linguistic symbols, but in other respects, the analysis discovers notable differences as to possibilities of interpolations and explanations, which, in the case of the novel, seem less evident. This difference is probably due to the fact that the objective information given in the encyclopaedia entry can be set against a background of commonly known frames and schemata, while the novel allows more freedom of interpretation, implying a somewhat less rigorous co-reference indexation.



As a summary of our presentation of various co-reference analyses, we shall briefly outline further research on the subject, as it appears in two, more recent volumes of *Officina Textologica*.

Volume 4 (2000), based on the typological investigations of *Volume 2*, discusses, in 7 articles, issues of a general scope in a theoretical approach. The volume starts with a survey, by GÁBOR TOLCSVAI NAGY, of the international specialist literature in text linguistics / textology exploring problems of co-reference and its relation to anaphor. EDIT DOBI elaborates a complex, semi-formal, semiotic-textological model of Chomskyan inspiration, proposed as a broader framework for the description of co-reference relations, a particular aspect of which, namely the role of thesaurus-like lexico-semantic explications in coreferential analysis, is presented by JÁNOS S. PETŐFI and EDIT DOBI. The last study in the volume, by ANDRÁS KERTÉSZ, examines the problem of ‘reference’ and ‘co-reference’ in the language philosophy of Frege, where ‘reference’ equals the referent (*Bedeutung*) of a linguistic expression, while ‘co-reference’ is explicated as the identity of referents between two linguistic expressions, whether they have the same sense (*Sinn*) or not. — Beside these theoretical studies, the volume contains concrete analyses of coreferential phenomena, in a highly rhetorical poem (LÓRÁNT BENCZE), and in an extract from *New Testament* (KÁROLY I. BODA—JUDIT PORKOLÁB), the latter article demonstrating a new model for computer-based text processing. Finally, a

practice-oriented study by ZSUZSA BENKES proposes a series of creative exercises for the analytical examination of co-reference relations.

Volume 12 (2005) marks a certain change in research orientation: while previous co-reference analyses worked with Hungarian texts, this time coreferential phenomena are examined in a polyglot contrastive approach, comparing Hungarian with French (3 articles), English (2 articles) and German (1 article). The analyses are either based on linguistic data obtained from various types of corpora, non-literary and literary texts (occasionally a text and its translation[s]) or on examples construed by the authors for the illustration of a coreferential phenomenon. The former method is used by ANDREA CSÚRY in her analysis of dialogue in French and Hungarian literary texts, by ISTVÁN CSÚRY, who uses a large corpus of political debates, also in French and Hungarian, by KÁROLY I. BODA—JUDIT PORKOLÁB in a comparison of a poem by T. S. Eliot and its two Hungarian translations, and by EDIT DOBI, who, in a German-Hungarian contrastive analysis, examines two extracts (with their translations) representing two text types: one written in specialist language, the other being a literary narrative. Finally, two studies investigate particular semantic-grammatical problems: co-reference relations in the conceptual structure of English modal auxiliaries and their Hungarian counterparts, by PÉTER PELYVÁS, and pronominal representation of co-reference in French and Hungarian by SÁNDOR KISS—FRANCISKA SKUTTA.

These three volumes of *Officina Textologica*, covering all main aspects of co-reference relations, will hopefully serve as a starting point for further theoretical research, and might as well be useful for educational purposes.

References

- Petőfi S. János 1997. *Egy poliglott szövegnyelvészeti-szövegtani kutatóprogram. Officina Textologica 1*. Debrecen: Kossuth Egyetemi Kiadó. 24–38.
- Petőfi S. János 1998. Koreferenciális kifejezések és koreferenciarelációk Példaszöveg: Mt. 9, 9-13. Máté meghívása. In: *Koreferáló elemek – koreferenciarelációk* = *Officina Textologica 2*. Debrecen: Kossuth Egyetemi Kiadó. 15–31.
- Petőfi S. János–Dobi Edit 1998. Utószó. In: *Koreferáló elemek – koreferenciarelációk* = *Officina Textologica 2*. Debrecen: Kossuth Egyetemi Kiadó. 238–261.