LOCAL PROPERTIES OF TEXTS

Jerzy Pogonowski

Adam Mickiewicz University, Poznań

We continue in this paper investigations into the properties of the mathematical model of linguistic analysis described in (Pogonowski 1978). Here we are interested especially in relations which hold between texts from any fixed language level. We will consider relations determined between whole texts as well as relations between fragments of these texts. First of all we must decide what we mean in a modeltheoretic interpretation of analyzed texts by a fragment of texts. It is worth-while to precede the formal definition with some explanatory remarks. Assume, for example, that we are considering immediate constituent analysis. Some finite tree (of immediate constituents) is associated with every sentence (of a given language). We can then investigate similarities (resp. distinctions) between these trees. As a consequence we can obtain some general theorems about the considered linguistic analysis. Another problem is: find the general (possible) form of the noun/verb phrase in a given language.

We assume that the reader is familiar with the notions of substructure (of a relational structure), monomorphism and isomorphism of relational structures. The definitions can be found in Sacks (1972).

According to (Pogonowski 1978) let S_1 be a set of relational structures corresponding to the family of analyzed texts from a given language level. We recall definitions of some special model-theoretic concepts. (It suffices for our purposes to relate these concepts to the set S_1).

Definition 1.

a) Let $\mathfrak{A}, \mathfrak{B} \in S_1$. We say that $\mathfrak{A}$ and $\mathfrak{B}$ are *locally isomorphic* if there exists a function f such that:

1. the domain of f (denoted dom (f)) is included in $|\mathfrak{A}|$
2. the range of f (denoted rng (f)) is included in $|\mathfrak{B}|$

3. the substructures of $\mathfrak{A}$ and $\mathfrak{B}$ with domains dom (f) and rng (f) respectively are isomorphic.

b) Let Φ be a set of formulae from L (Ω_i). By a *Φ-morphism* between $\mathfrak{A}$ and $\mathfrak{B}$ we mean any monomorphism from $\mathfrak{A}$ to $\mathfrak{B}$ which preserves the validity of all formulae from Φ.

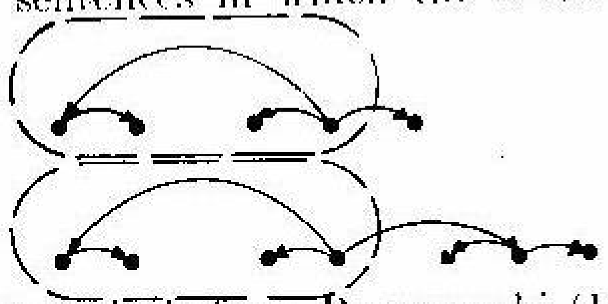
c) By a *local Φ-morphism* between $\mathfrak{A}$ and $\mathfrak{B}$ we mean a function f such that

1. dom (f) $\subseteq |\mathfrak{A}|$ rng (f) $\subseteq |\mathfrak{B}|$
2. f is a Φ-morphism between substructures of $\mathfrak{A}$ and $\mathfrak{B}$ with domains dom (f) and rng (f) respectively.

The linguistic sense of the above definitionsis illustrated by the following example.

Example 1.

a) The local isomorphism between structures associated with analysed texts corresponds to the "structural indistinguishability" of these texts. We have, for example, the local isomorphism between the following structures which are associated with sentences in which the relation of concord is marked.



b) Let us consider Ω-analysis from Pogonowski (1978), example 1. Let φ be the following sentence from L (Ω_2):

$$\bigvee_{v_1} \bigwedge_{v_2} (P_1(v_1, v_2) \rightarrow P_2(v_1, v_2))$$

Then local $\{\varphi\}$-isomorphism holds between these structures from S_2 which have isomorphic substructures satisfying the sentence φ. This sentence "says" that there exists a word such that any word which follows it is governed by it:

Before a more exact examination of the relations between texts and their fragments, we will recall a theorem well known in model theory This theorem can also be applied to our model of linguistic analysis.

Theorem 1.

For any $\mathfrak{A}, \mathfrak{B} \in S_i$ the following two conditions are equivalent:

1. $\mathfrak{A}$ and $\mathfrak{B}$ are elementarily equivalent (we denote this fact by $\mathfrak{A} \equiv \mathfrak{B}$)
2. $\mathfrak{A}$ and $\mathfrak{B}$ are isomorphic (we denote this fact by $\mathfrak{A} \simeq \mathfrak{B}$)

In model theory the existence of isomorphism between relational structures $\mathfrak{A}$ and $\mathfrak{B}$ means that these structures are structurally indistinguishable. On the other hand elementary equivalence between $\mathfrak{A}$ and $\mathfrak{B}$ means that $\mathfrak{A}$ and $\mathfrak{B}$ have the same first-order proprieties (that $\mathfrak{A}$ and $\mathfrak{B}$ satisfy the same

sentences from L (Ω_i)). For $\mathfrak{A}, \mathfrak{B} \in S_i$ denote by t$\mathfrak{A}$ and t$\mathfrak{B}$ the texts (from i-th language level) which are associated with $\mathfrak{A}$ and $\mathfrak{B}$ respectively. Then the implication

$$\text{if } \mathfrak{A} \equiv \mathfrak{B} \text{ then } \mathfrak{A} \simeq \mathfrak{B}$$

can be explained in the following way:
If everything that is true for text t$\mathfrak{A}$ is true for text t$\mathfrak{B}$ and conversely, then the texts t$\mathfrak{A}$ and t$\mathfrak{B}$ are structurally indistinguishable.
From theorem 1 immediately follows:

Corollary 1.

For any $\mathfrak{A}, \mathfrak{B} \in S_i$, if $\mathfrak{A} \equiv \mathfrak{B}$, then the sets $|\mathfrak{A}|$ and $|\mathfrak{B}|$ have the same number of elements.

This corollary shows us why it is better to investigate (in our model of linguistic analysis) local morphisms between relational structures instead of isomorphisms.

We define for any $\mathfrak{A} \in S_i$:

$$P(\mathfrak{A}) = \{\mathfrak{B} : \mathfrak{B} \text{ is a nonempty substructure of } \mathfrak{A}\}$$

Let $P(S_i) = \cup\{P(\mathfrak{A}) : \mathfrak{A} \in S_i\}$.

Finally, let $IP(\mathfrak{A}) = \{\mathfrak{B} \in P(S_i)$: there exists $\mathfrak{C} \in P(\mathfrak{A})$ such that $\mathfrak{B}$ is isomorphic with $\mathfrak{C}\}$.

In our model-theoretic interpretation of linguistic analysis the set P $(\mathfrak{A})$ corresponds to the family of all fragments of the text associated with $\mathfrak{A}$. Hence P (S_i) corresponds to the family of all fragments of all texts from the i-th language level.

Finally, for $\mathfrak{A} \in S_i$ the set IP $(\mathfrak{A})$ corresponds to the family of all fragments (of texts from i-th language level) which are structurally indistinguishable from fragments of the text associated with $\mathfrak{A}$.

The following theorem shows the properties of local isomorphism.

Theorem 2.

Let $\mathfrak{A}, \mathfrak{B} \in S_i$. The relation ∇ defined by $\mathfrak{A} \nabla \mathfrak{B}$ if and only if $IP(\mathfrak{A}) \cap IP(\mathfrak{B}) \neq 0$ is a tolerance relation on S_i.

We recall that by a tolerance relation we mean a binary relation which is reflexive and symmetric. It is easy to see that $\mathfrak{A} \nabla \mathfrak{B}$ is equivalent to the existence of local isomorphism between $\mathfrak{A}$ and $\mathfrak{B}$. However, one thing ought to be stressed. Namely, if Ω_i contains few predicates, then it may often happen that the relation ∇ holds between any two structures from S_i. For this reason the investigations of the relation ∇ become nontrivial only for "rich" linguistic analysis (rich with respect to the number of considered relations between language entities). We will show, in another paper (see Pogonowski 1979), how one can use ∇ to construct global (topological) structures in the set S_i.

We recall that S_i corresponds to the i-th language level. It is worth pointing out here that relation ∇ is an example of the relation determined by connections between fragments of texts. It seems natural to call relations of this kind local properties of texts. It is clear that the investigation of these phenomena is of fundamental importance to linguistics. Local properties of texts serve as a basis for resolving problems connected with models of translation, the typology of languages, descriptions of sentence patterns and so on.

In general, we must consider not only the relation ∇ but also some other relations corresponding to other kinds of text similarity.

The following theorem is an example of the classification of texts with respect to their local properties.

Theorem 3.

Let $\mathfrak{A}, \mathfrak{B} \in S_l$. Define relation ∇^* by:

$\mathfrak{A} \nabla^* \mathfrak{B}$ if and only if there exists a sequence $\mathfrak{A}_1, \mathfrak{A}_2, \ldots, \mathfrak{A}_n$ of structures from S_l such that $\mathfrak{A} = \mathfrak{A}_1$, $\mathfrak{B} = \mathfrak{A}_n$ and for all m: if $1 \leqslant m < n$ then $\mathfrak{A}_m \nabla \mathfrak{A}_{m+1}$. Then ∇^* is an equivalence relation on S_l.

Our preceding investigations have been concerned with text relations defined according to similarities in text fragments. One can pose another question in some sense parallel to the previous one. Namely, given a fixed fragment of some text, how can we determine the family of all texts which contain a fragment structurally indistinguishable from the given fragment? Formally, this problem can be written in the following way: Let $\mathfrak{B} \in P(S_l)$. What can we say about the set $N(\mathfrak{B}) = \{\mathfrak{A} \in S_l : \mathfrak{B} \in IP(\mathfrak{A})\}$? From the definition of the sets $N(\mathfrak{B})$ some corollaries follow immediately.

Corollary 2.

If $\mathfrak{B}_1, \mathfrak{B}_2 \in P(S_l)$ and $\mathfrak{B}_1 \simeq \mathfrak{B}_2$ then $N(\mathfrak{B}_1) = N(\mathfrak{B}_2)$.

Corollary 3.

If $\mathfrak{B}_1, \mathfrak{B}_2 \in P(S_l)$ and there exists a monomorphism from $\mathfrak{B}_1$ to $\mathfrak{B}_2$, then $N(\mathfrak{B}_2) \subseteq N(\mathfrak{B}_1)$.

Corollary 4.

The relation $\sim_N$ defined by

$$\mathfrak{B}_1 \sim_N \mathfrak{B}_2 \text{ if and only if } N(\mathfrak{B}_1) = N(\mathfrak{B}_2)$$

is an equivalence relation on $P(S_l)$.

The family $\{N(\mathfrak{B}) : \mathfrak{B} \in P(S_l)\}$ is a cover of the set S_l, i.e.:

$$\bigcup_{\mathfrak{B} \in P(S_l)} N(\mathfrak{B}) = S_l$$

However, there exists a minimal cover of S_l with the sets $N(\mathfrak{B})$. To obtain this cover it suffices to choose one structure $\mathfrak{B}_n$ from each equivalence class of the relation $\sim_N$: $\bigcup_n N(\mathfrak{B}_n) = S_l$.

A similar question can be asked for $P(S_l)$. Namely, given $P(S_l) = \bigcup_{\mathfrak{A} \in S_l} IP(\mathfrak{A})$, can we look for minimal covers of $P(S_l)$?

REFERENCES

Ehrenfeucht, A. 1961. "An application of games to the completeness problem for formalized theories". *Fundamenta Mathematicae 49*. 129–141.
Pogonowski, J. 1978. "Mathematical model of linguistic analysis". *SAP 10*. 123-130
Pogonowski, J. 1979. "Global properties of language levels". *SAP 11*.
Sacks, G. 1972. *Saturated model theory*. Mass.: W. A. Benjamin.