GLOBAL PROPERTIES OF LANGUAGE LEVELS

JERZY POGONOWSKI

Adam Mickiewicz University, Poznań

This paper is a continuation of investigations contained in Pogonowski (1979a) and (1979b). It is devoted to a description of some algebraic and topological structures on language levels. We will show that local properties of analyzed texts can be used for introducing global structures on the whole language level. All terminology and notations used here are defined and explained in Pogonowski (1979a, 1979b).

We recall that a relational structure is associated with every analyzed text in our model of linguistic analysis. Consequently, a family of relational structures (of the same type) is associated with each language level. All our considerations are true for any arbitrary fixed language level. Let S_i denote a family of relational structures which corresponds to the i-th language level. We will consider two types of structures on S_i: *A*. algebraic, *B*. topological. We assume that the reader is familiar with elementary notions from model theory, universal algebra and general topology.

A. Algebraic structures on language levels

In Pogonowski (1979a) we considered a simple algebraic structure established on a particular language level. We recall it briefly here. Let our Ω-analysis be i+1-linear. Assume that h is an equivalence relation on S_i defined without the use of predicates from Ω_{i+1}. Then there exists a mapping g_i from the set S_{i+1} in free semigroup F_h generated by the family of h-equivalence classes $S_{i/h}$. However, the algebraic structure of $g_i[S_{i+1}]$ induced from F_h is rather simple. The inclusion $g_i[S_{i+1}] \subseteq F_h$ means that texts from the i+1-st language level are linear combinations of texts from the i-th language level. In this case we do not say anything about linguistic relations between texts.

We introduce a richer algebraic structure on S_i, using some notions defined in Pogonowski (1979b). Recall the needed definitions:

for $\mathfrak{A} \in S_i$ define $\mathrm{IP}(\mathfrak{A}) = \{\mathfrak{B} \in P(S_i) : \mathfrak{B}$ is isomorphic to some substructure of $\mathfrak{A}\}$

for $\mathfrak{B} \in P(S_i)$ define $N(\mathfrak{B}) = \{\mathfrak{A} \in S_i : \mathfrak{B} \in \mathrm{IP}(\mathfrak{A})\}$.

Here $P(S_i) = \bigcup \{P(\mathfrak{A}) : \mathfrak{A} \in S_i\}$, and $P(\mathfrak{A})$ denotes the family of all non-empty substructures of $\mathfrak{A}$. The relation $\mathfrak{B} \in \mathrm{IP}(\mathfrak{A})$ induces a pair of functions between $\mathscr{P}(S_i)$ and $\mathscr{P}(P(S_i))$, (families of all subsets of S_i and $P(S_i)$, respectively). Namely, we define:

for $K \subseteq S_i$ $\quad K^* = \bigcap_{\mathfrak{A} \in K} \mathrm{IP}(\mathfrak{A})$ $\quad$ (then $K^* \subseteq P(S_i)$)

for $L \subseteq P(S_i)$ $\quad L^* = \bigcap_{\mathfrak{B} \in L} N(\mathfrak{B})$ $\quad$ (then $L^* \subseteq S_i$)

Theorem 1.

Let $K \subseteq S_i$, $L \subseteq P(S_i)$. The pair of functions between sets $\mathscr{P}(S_i)$ and $\mathscr{P}(P(S_i))$ defined by

$$K \to K^*, \quad L \to L^*$$

is a *Galois connection*.

For a proof of the theorem one must check that the following conditions are true:

1. if $K_1 \subseteq K_2$ then $K_2^* \subseteq K_1^*$ (for $K_1, K_2 \subseteq S_i$)
2. if $L_1 \subseteq L_2$ then $L_2^* \subseteq L_1^*$ (for $L_1, L_2 \subseteq P(S_i)$)
3. $K \subseteq K^{**}$ (for $K \subseteq S_i$)
4. $L \subseteq L^{**}$ (for $L \subseteq P(S_i)$)

We leave the proof, which is easy, to the reader. Here K^{**} denotes $(K^*)^*$ of course.

Any Galois connection determines some closure operator. Namely, we define operator $D: \mathscr{P}(S_i) \to \mathscr{P}(S_i)$ by $D(K) = K^{**}$. Then D has the following properties:

1. if $K_1 \subseteq K_2$ then $D(K_1) \subseteq D(K_2)$
2. $K \subseteq D(K)$
3. $D(D(K)) = D(K)$

Example 1.

Let S_i correspond to the family of all sentences of some fixed language. Assume that syntactic categories (tense, voice, mood, etc.) are determined by the occurrence of a phrase with a strictly defined frame. Consider some set of sentences with only one common phrase (with respect to its structure) — say, a phrase determining active voice. Let K be the family of relational structures from S_i which corresponds to this set. Then D(K) is the family of structures which corresponds to the set of *all* active sentences. Similarly for the other categories mentioned above. It is interesting to determine when D is the

algebraical closure operator, i.e. whether the following condition holds:

$$(*)\ \begin{cases} \text{if } \mathfrak{A} \in D(K),\ K \subseteq S_i \text{ then there exists a finite set } K_0 \subseteq K \text{ such that} \\ \mathfrak{A} \in D(K_0). \end{cases}$$

The validity of (*) depends of course on the set S_i. Let us consider the simplest case, when S_i is finite (only a finite number of texts is considered). Then D is the algebraic closure operator, because in this case each subset of S_i is finite. Before explaining this fact and its importance for linguistic analysis we must recall some notions from universal algebra.

1. Any closure operator D on the set A determines a system $\mathscr{C}$ of closed sets in A, i.e. a family $\mathscr{C}$ of subsets of A which is closed with respect to set-theoretical intersection:

$$\text{for any } \mathscr{A} \subseteq \mathscr{C} : \bigcap \mathscr{A} \in \mathscr{C}$$

The required family $\mathscr{C}$ is defined by:

$$\mathscr{C} = \{B \subseteq A : D(B) = B\}$$

The family $\mathscr{C}$ is called an *algebraic system of closed sets* if the corresponding closure operator D is algebraic.

2. (Schmidt's theorem). Assume that in the set A some algebraic system $\mathscr{C}$ of closed sets is determined. Then one can introduce the structure of *algebra* into A. In addition, the family $\mathscr{C}$ is the family of all sub-algebras of this algebra.

We use the above facts to show when one can associate some interesting algebraic structure with the i-th language level. The following theorem is a simple consequence of Schmidt's theorem:

Theorem 2.

Assume that the set S_i is finite. Let

$$\mathscr{C} = \{K \subseteq S_i : D(K) = K\}$$

where D is the closure operator defined by the Galois connection from theorem 1 $D(K) = K^{**}$. Then one can define operations $o_1, o_2, \ldots, o_{n_i}$ such that $\langle S_i, o_1, o_2, \ldots, o_{n_i} \rangle$ will from an *algebra*. All sub-algebras of this algebra are elements of the family $\mathscr{C}$. The linguistic sense of this theorem is illustrated by the following example:

Example 2.

Let us make the same assumptions as in example 1. Denote by $\mathscr{C}$ the system of closed sets determined by the closure operator D from example 1. Then the family $\mathscr{C}$ will contain sets of structures which correspond to:

the family of al' active sentences; and
the family of all passive sentences, among others.

One can obtain, in an easy way, analogons of theorems 1, 2 for the family $P(S_i)$. It would be interesting to construct the algebra mentioned in theorem 2 in the case of concrete, sufficiently rich, but not too complicated Ω-analysis.

B. Topological structures on language levels

Let $\mathfrak{M}_{\Omega_i}$ denote the class of all relational structures of the type Ω_i. It is well known (see Bell and Slomson 1969, for example) that $\mathfrak{M}_{\Omega_i}$ is a compact topological space with a basis consisting of closed-open sets. Because $S_i \subseteq \mathfrak{M}_{\Omega_i}$ we can introduce in S_i a subspace topology. We are interested in these topological properties of S_i from which follow corollaries important for linguistic analysis. We start with a short presentation of some facts concerning subspace topology on S_i induced from $\mathfrak{M}_{\Omega_i}$. If φ is a sentence from $L(\Omega_i)$ then denote:

$$Mod(\varphi) = \{\mathfrak{A} \in \mathfrak{M}_{\Omega_i} : \varphi \text{ is valid in } \mathfrak{A}\}$$

If Φ is a set of sentences from $L(\Omega_i)$ denote

$$Mod(\Phi) = \bigcap_{\varphi \in \Phi} Mod(\varphi).$$

The family $\mathscr{B}_i$ defined by

$\mathscr{B}_i = \{K \subseteq S_i$: there exists a sentence φ from $L(\Omega_i)$ such that for all $\mathfrak{A} \in K$, φ is valid in $\mathfrak{A}\}$ is a basis of subspace topology on S_i. Hence open sets in this topology are of the form

$$K_n = \bigcup_n K_n \quad \text{where } K_n \in \mathscr{B}_i$$

It is clear for any sentence φ from $L(\Omega_i)$

$$S_i - Mod(\varphi) = S_i \cap Mod(\neg \varphi)$$

From this follows that closed sets in subspace topology are of the form

$$K_n = \bigcap_n K_n \quad \text{where } K_n \in \mathscr{B}_i$$

All elements of the family $\mathscr{B}_i$ are simultaneously open and closed.

Subspace topology on S_i is compact if and only if S_i is a closed subset of $\mathfrak{M}_{\Omega_i}$. On the other hand, S_i is closed (in $\mathfrak{M}_{\Omega_i}$) if and only if there exists a set Φ of sentences from $L(\Omega_i)$ such that $S_i = Mod(\Phi)$. In terms of linguistic analysis, the existence of the set Φ such that $S_i = Mod(\Phi)$ means that the i-th language level can be described axiomatically. A necessary and sufficient condition for the existence of a set Φ with the above property is the following equality:

$$S_i = Mod(Th(S_i))$$

(here $Th(S_i) = \bigcap_{\mathfrak{A} \in S_i} Th(\mathfrak{A})$ and $Th(\mathfrak{A})$ is the set of all sentences from $L(\Omega_i)$, valid in $\mathfrak{A}$)

Hence, we see that the possibility of the existence of an axiomatic description

in $L(\Omega_i)$ of the i-th language level depends on our assumptions on the set S_i. For example, let the family S_i correspond to the level of all sentences of a fixed language. Assume that we admit the possibility of forming arbitrarily long (but still finite!) sentences. In this case the family S_i will contain finite structures of arbitrarily high cardinal. It is easy to show (using ultraproduct construction) that in such a case there *does not* exist a set Φ of sentences from $L(\Omega_i)$ such that $S_i = Mod(\Phi)$. In terms of linguistic analysis: if we admit arbitrarily long sentences, then the property "to be a sentence" cannot be described axiomatically in the language of linguistic analysis.

We will investigate some other topological structures on S_i. We will show that the relation ∇, defined in Pogonowski (1979b) by $\mathfrak{A} \nabla \mathfrak{B}$ if and only if $IP(\mathfrak{A}) \cap IP(\mathfrak{B}) \neq 0$ determines some topologies on S_i. From the definition of the relation it follows that these topologies are connected with local properties of texts. In Pogonowski (1979b) we proved that ∇ is a tolerance relation i.e. it satisfies the following conditions:

$$\mathfrak{A} \nabla \mathfrak{A} \qquad \text{(reflexivity)}$$
$$\mathfrak{A} \nabla \mathfrak{B} \rightarrow \mathfrak{B} \nabla \mathfrak{A} \quad \text{(symmetry)}$$

Thus (S_i, ∇) is a tolerance space. Before investigating the topological structures determined by ∇ we recall some facts about tolerance spaces. Let (X, τ) be any (fixed) tolerance space. We say that $A \subseteq X$ is a *preclass* if for all $x, y \in A$: : $x \tau y$. Maximal (with respect to inclusion) preclasses are called *classes*. Each preclass is contained in some class. Denote by H_X the family of all classes in (X, τ). The family H_X is a cover of X. For any $x \in X$ let $H_X(x) = \{A \in H_X : x \in A\}$. Define

$$x \tau^+ y \equiv \bigwedge_{z \in X} (x \tau z \equiv y \tau z).$$

Then τ^+ is an equivalence relation on S_i. The set of all τ^+-equivalence classes is denoted by $\hat{X}$. Elements of the set $\hat{X}$ are called *kernels*. We denote by $J(x)$ the kernel which contains x. Thus $\hat{X} = \{J(x) : x \in X\}$. A tolerance space (X, τ) is called:

1. *simple*, if $J(x) = \{x\}$ for all $x \in X$
2. *regular*, if $\bigwedge_{x \in X} J(x) = \bigcap H_X(x)$

A family $\mathfrak{B} \subseteq H_X$ is called a *basis* of tolerance space (X, τ) if the following conditions hold:

1. if $x \tau y$ then there exists $A \in \mathfrak{B}$ such that $x, y \in A$
2. $\mathfrak{B}$ is a minimal family satisfying 1

For every tolerance space there exists at least one basis.

Now we come back to our model of linguistic analysis. First of all, we show the interpretation of notions introduced above in the case of the considered tolerance space (S_i, ∇). Any preclass in (S_i, ∇) corresponds to the family of texts which are all similar in the sense that any two texts from this family have a common phrase with respect to structure. If $A \subseteq S_i$ is a class then the corresponding family of texts has the following property: there is no text in $S_i - A$ similar (in the sense mentioned above) to all texts from A. Every kernel $J(\mathfrak{A}) \in \hat{S}_i$ consists of texts which have exactly the same common (with respect to structure) phrases. It is interesting to find (and describe) a basis $\mathfrak{B}$ of (S_i, ∇) for a concrete fixed Ω-analysis. The solution of this problem may be useful for language teaching: any basis $\mathfrak{B}$ of (S_i, ∇) is the most economical description of the structural degree of complexity of the i-th language level and knowing a basis $\mathfrak{B}$ we can reconstruct all structural similarities between texts. We can obtain several topologies on S_i using the relation ∇:

1. For any $\mathfrak{A} \in S_i$ denote $\nabla(\mathfrak{A}) = \{\mathfrak{B} \in S_i : \mathfrak{A} \nabla \mathfrak{B}\}$.
Let $\mathfrak{I}^* = K \subseteq S_i$: for every $\mathfrak{A} \in K$, $\nabla(\mathfrak{A}) \subseteq K\}$.

Lemma 1.

$\mathfrak{I}^*$ is a topology on S_i

Every open set $K \in \mathfrak{I}^*$ corresponds to the family of texts which, for every element, contains all elements similar to this element. Topology $\mathfrak{I}^*$ has the following properties:

Theorem 3. (Hartnett)

1. Intersection of any family $K \subseteq \mathfrak{I}^*$ belongs to $\mathfrak{I}^*$
2. There exist minimal open sets in $\mathfrak{I}^*$
3. Denote by $Tr(\nabla)$ the transitive closure of ∇, i.e. the least transitive relation containing ∇. Then $Tr(\nabla)$ is of course an equivalence relation. The family $S_i/_{Tr(\nabla)}$ of all $Tr(\nabla)$-equivalence classes is a basis for the topology $\mathfrak{I}^*$. For any other basis $\mathscr{B}$ the following inclusion holds:

$$S_i/_{Tr(\nabla)} \subseteq \mathscr{B}$$

However, the topology $\mathfrak{I}^*$ seems to be a rather rough structure. We need more subtle tools for description of the structural similarities (respectively distinctions) between texts.

2. For any $\mathfrak{A} \in S_i$ and $K_1, \ldots, K_n \in H_{S_i}(\mathfrak{A})$ define:

$$U(\mathfrak{A}, K_1, \ldots, K_n) = \bigcap_1 K_i$$

$$\mathscr{B}(\mathfrak{A}) = \{K \subseteq S_i: \text{there exists } U(\mathfrak{A}, K_1, \ldots, K_n) \text{ such that } U(\mathfrak{A}, K_1, \ldots, K_n) \subseteq K\}.$$

It is easy to prove the following lemma:

Lemma 2.

a) for every $\mathfrak{A} \in S_i$, $\mathscr{B}(\mathfrak{A}) \neq 0$
b) for all $U \in \mathscr{B}(\mathfrak{A})$, $\mathfrak{A} \in U$
c) if $\mathfrak{A} \in U \in \mathscr{B}(\mathfrak{B})$ then there exists $V \in \mathscr{B}(\mathfrak{A})$ such that $V \subseteq U$
d) if $V_1, V_2 \in \mathscr{B}(\mathfrak{A})$ then there exists $V \in \mathscr{B}(\mathfrak{A})$ such that $V \subseteq V_1 \cap V_2$ [1*]

From lemma 2 it follows that for every $\mathfrak{A} \in S_i$ the family $\mathscr{B}(\mathfrak{A})$ is a fundamental system of neighbourhoods of $\mathfrak{A}$. Hence the family $\{\mathscr{B}(\mathfrak{A}) : \mathfrak{A} \in S_i\}$ determines some topology $\mathfrak{I}$ on S_i:

$\mathfrak{I} = \{K \subseteq S_i : K$ is a sum of some subfamily of the family $\bigcup_{\mathfrak{A} \in S_i} \mathscr{B}(\mathfrak{A})\}$.

Properties of the space $(S_i, \mathfrak{I})$ are described by the following theorem.

Theorem 4. (Shreider)

a) for any $\mathfrak{A}, \mathfrak{B} \in S_i$: $\mathfrak{A} \nabla \mathfrak{B}$ if and only if there exists $U(\mathfrak{A}, K)$ such that $\mathfrak{B} \in U(\mathfrak{A}, K)$
b) if (S_i, ∇) is simple, then $(S_i, \mathfrak{I})$ is T_0-space
c) $(S_i, \mathfrak{I})$ is Hausdorff space if and only if (S_i, ∇) is simple and regular.

The topology $\mathfrak{I}$ is richer than $\mathfrak{I}^*$. Indeed, the identity map id : $(S_i, \mathfrak{I}) \to (S_i, \mathfrak{I}^*)$ is continuous (but id is not a homeomorphism!). Any two elements $\mathfrak{A}, \mathfrak{B} \in S_i$ with the same kernel cannot be separated by an open set (we say that in this case $\mathfrak{A}$ and $\mathfrak{B}$ are arbitrarily near).

3. Some quasi-topological structure is connected with every tolerance space. Namely, we define for every $A \subseteq S_i$

$$d(A) = \{\mathfrak{B} \in S_i : \text{there exists } \mathfrak{A} \in A \text{ such that } \mathfrak{A} \nabla \mathfrak{B}\}.$$

Of course $d(A) = \bigcup_{\mathfrak{A} \in A} \nabla(\mathfrak{A})$. The operator $d : \mathscr{P}(S_i) \to \mathscr{P}(S_i)$ has the following properties:

1. $d(0) = 0$
2. $d(A \cup B) = d(A) \cup d(B)$
3. $A \subseteq d(A)$
4. $d(A) \subseteq d(d(A))$

In general, inclusion in 4) cannot be replaced by equality. For this reason d is not a topological closure operator (in Kuratowski's sense). The quasi-topological structure introduced in S_i by d is very interesting. It is also useful in applications, especially in the case when the set S_i is finite (i.e. if we consider only a finite number of texts). If S_i is finite, then every topology on S_i which

* Z. Saloni pointed out that lemma 2c (which is due to Schreider) does not hold. However, theorem 4 holds, if we define $\mathfrak{I}$ to be the topology determined by the sub-base H_{S_i}.

is at least T_1-topology is automatically discrete and so useless for any applications. For this reason it seems to be better to investigate more general (than classical) topological structures in finite sets. It is reasonable to call the structure determined by the operator d *many-staged topology* (shortly: ms-topology). One can define analogons of some classical topological notions for ms-topologies (closure, interior, boundary). ms-topology on (S_1, ∇) will be investigated in more detail in one of our later articles.

Let us finish this paper with some general remarks. Topologies on S_1 show the geometrical structure of language. This structure can be further investigated with the help of elementary homological algebra and algebraic topology. One can obtain results similar to the above using the sets $N(\mathfrak{B})$, $(\mathfrak{B} \in P(S_1))$ and the relation $\Delta : \mathfrak{A} \Delta \mathfrak{B} \equiv N(\mathfrak{A}) \cap N(\mathfrak{B}) \neq 0$. The linguistic counterpart of this relation (and constructions connected with it) is quite clear. Further, it is possible to show correspondence between the considered algebraic and topological structure on language levels (Galois connection and tolerances ∇, Δ). Topological invariants obtained in this way can be used for developing a mathematical typology of languages. It is worth pointing out that the global structures on language levels considered by us here are determined by local properties of texts. It is of course possible to investigate another (for example paradigmatic in the sense of Pogonowski 1979a) relation on S_1 and obtain different kinds of global structures.

REFERENCES

Bell, J. L. and A. B. Slomson. 1969. *Models and ultraproducts*. Amsterdam: North Holland.
Cohn, P. 1965. *Universal algebra*. New York: Harper and Row.
Kelley, J. L. 1975. *General topology*. New York — Heidelberg — Berlin: Springer Verlag.
Pogonowski, J. 1979a. "Mathematical model of linguistic analysis". *SAP 10*. 123-130.
Pogonowski, J. 1979b. "Local properties of texts". *SAP 10*. 131-135.