Maciej Piasecki, Computer Science Department, Wrocław University of Technology (piasecki@ci.pwr.wroc.pl)

Semantic and Syntactic Aspects of Reference and Quantification in Polish Noun Phrase

1. Introduction

Most of the works on formal description of NP meaning is based on English examples, where the representation of the meaning of determiner constructs the structure of representation for the complete NP. Traditionally, the determiner phrase is perceived as being the source of many important aspects of NP meaning like: quantification, anaphor, reference, presupposition, genericness. It can be argued (and it will be later in this paper) that the same aspects can be identified for Polish NPs, as well. However, having no clearly distinguished syntactic category in Polish, similar to the determiner (at least traditionally), a question is emerging: where to locate these aspects in the formal representation of the syntactic-semantic structure of Polish NP? An attempt to answer this question is the main goal of this paper.
We will start with recapitulation of some basic informal assumption concerning the meaning of NP introduced in Piasecki (2002). Next, a logical language of semantic representation structures called *Self-organising Logic of Structures* (*SLS*), formalising these assumptions, will be introduced. Instead of direct introduction of complicated *SLS* semantics, detailed informal descriptions of the meaning of *SLS* formulas will be given. Moreover, graphical notation for *SLS* formulas will be defined. In the second part of the paper, starting with short overview of some selected works on the description of Polish NP, *SLS* expressions will applied to construct, in a compositional way, the meaning representation of broad range of Polish NPs. Finally, some interesting possible extensions of the approach will be discussed.

2. NP Meaning – Basic Assumptions

All assumptions underlying the formal model of NP meaning, being utilised here, have been presented in Piasecki (2002b). Avoiding repeating them in details[1], only the list of the features describing the mutually independent aspects of NP meaning is recapitulated below:

1. *Genericity* with two values: *general* and *specific* - expressing whether *NP represents an object* in the case of *specific* value, or *NP represents a notion* (more or less precisely identified) when the value is *general*.
2. *Referentiality* with two values: *referential* and *attributive* – where the values of referentiality distinguish between a situation in which *the hearer is expected to establish a certain relation between NP and the identified entity* (referential NP) and a situation in which *the hearer is informed about the potential existence of an entity described by the features encoded by the descriptive content of NP* (attributive NP).
3. *Presupposition*, (here only existential presupposition) where the value is a constrain of the number of possible referents present in the context of interpretation of discourse.
4. *Quantification*, with a complex value being a pair of values of sub-features: *generalised quantifier* and *distributivity*; both will be fully explained in the next section.

It is worth to emphasise that value *attributive* is used here instead *potential* used formerly in Piasecki (2002b), and that "*attributive"* is used here in slightly different way than in literature, e.g. in Kamp & Reyle (1993), where it is often the name of the use of NPs in object position

[1] All of them will be implicitly reintroduced during the definition of the formal language below.

of predicative sentence with *to be*. Reference is limited in this paper only to the reference to some objects, excluding the other kinds (e.g. reference to: time periods, situations, properties etc.) from detailed considerations. Moreover, the range of NPs being described here will be restricted only to NPs denoting countable objects. We will not regard here NPs denoting masses, substances etc.

3. *Self-organising Logic of Structures* – a Base for Description of Polish NP Meaning.

Such aspects of the NP meaning like anaphoric link, referential status, presupposition etc. can be only fully analysed in the broader perspective of the discourse meaning analyse. *Discourse Representation Theory* (*DRT*), Kamp & Reyle (1993), has become a widely used standard of formal description of the discourse semantics and started the development of the *dynamic semantics*. Because standard *DRT* is in some points non-compositional, many different compositional versions of *DRT* have been proposed, see Eijck & Kamp (1997) for brief overview. However, most of them share the same property i.e. they strongly depend on previous indexing of the syntactic tree identifying anaphoric links. Next, most compositional 'DRTs' interpret anaphoric expressions non-compositionally by assigning to them the discourse referent with the same name as to the antecedent. Final shape of DRS of the whole discourse is 'encoded into' all small parts from which we compositionally build the whole. Moreover, mostly compositional DRTs pays a little attention to presupposition and reference. A logical language, called *Self-organising Logic of Structures* (*SLS*) Piasecki (2002a), being presented here, originates from the attempt to create the theory of fully compositional and possibly complete description of NP meaning perceived from the discourse perspective. The important goal was to deliver a formal tool enabling compositional representation of nominal anaphor and reference. *SLS* is not a theory of computational anaphor resolution. It formulates only necessary syntactic-semantic conditions, which must be preserved by discourse representation.
SLS operators are defined as kind of short-cuts expressed in a variant of standard many-sorted type logic. However, in order to simplify the linguistic applications of *SLS* a graphical representation of a subset[2] of *SLS* expressions has been introduced. In the rest of the paper, we will rest on it, giving only some necessary general information about formal details.
SLS preserves a dynamic notion of meaning as a change introduced by utterance to the context of interpretation. In order to represent reference (and presupposition) the context must be broaden with some elements of the knowledge possessed by the hearer and concerning the surrounding world. Such context will be called here an *information state* (of the hearer). The information state is represented in *SLS* by objects of the compound (logical) type called *state* represented by the letter *s*. The basic *SLS* formulas used for the representation of many expressions (including all compound) of the natural language are relations on pairs of states of type (*s*(*st*)) and are called *dynamic formulas*. Dynamic formulas change an input state into an output state and their meaning can be equalled to the change being introduced by them. A dynamic formula is true in some *SLS* model (universe + constants interpretation) if this relation is not empty.
The notion of the state is based on the metaphor of 'computer memory'. Similarly to other versions of *DRT*, information about the subsequent objects, introduced into the domain of discourse by the use of the following NPs, is 'stored' in *discourse referents* - elements of type *m*. Discourse referents represent memory cells and the denotation of *m*, written $\mathbf{D}_{\mathrm{m}}$ is a infinite set with a total order forming a linear structure of the 'memory'. Because the context includes also information originating from the outside of the discourse, 'memory cells' (the discourse

[2] The subset utilised in the compositional analyse of the discourse meaning.

referents) can contain information about objects known to the hearer from the world.
The state is a structure of the compound type: ($m \bullet ((m \bullet m)t) \bullet (m(et))$). The state, graphically represented on the Fig. 1, consist of the three parts:

1. *a set of activated referents* given by the single, distinguished referent, pointing to the top of the 'area of the memory' used until a certain moment in the interpretation of the discourse; all referents below and equal the most recently used referent are called activated referents and are accessible for the anaphoric and referential links;
2. *a set of pairs of referents* – describes links between referents, the first referent in each pair, being an antecedent, can be only a referent activated in the given state;
3. *a valuation function* – according to plurality and quantification of NPs, assign sets of *objects* from the *SLS* universe to the activated referents; if there are some links for a referent then at least in the case of one antecedent the assigned value must be identical for both referents (it is guaranteed by the *SLS* axiom of *Indeterministic Interpretation of Reference*).

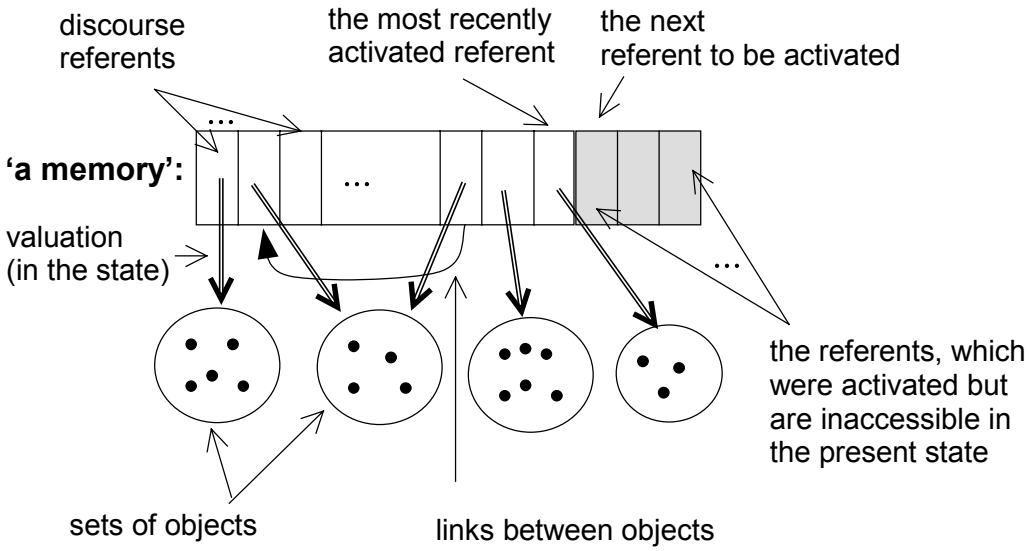


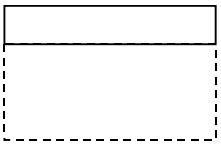
Fig. 1 'A memory': the outline of the structure of the information state.

The usage of the same mechanism of a link for the representation of the reference and anaphor in SLS does not mean that anaphor is being identified with co-reference. The values of referents are assigned by the mechanism of *existential import*, Kamp & Reyle (1993), included in the meaning of *referent activation operator*. The objects known to the hearer are represented by referents being activated in the initial, non-empty, state of the discourse interpretation.
Having the state defined, the *SLS* operators can be informally introduced, together with the definitions of the graphical representation of basic *SLS* expressions utilising them.

Fig. 2 Examples of noun predicative expression.

Each NP identifies a *class* of entities fulfilling the constraints expressed by a descriptive content of the NP. The descriptive content is represented by a *noun predicative expression* of type (*et*), where *e* and *t* are standard types of, respectively, *entities* and *truth values*. The predicative expression can be a single predicate, e.g. *mężczyzna* or conjunction of predicates *e.g. przystojny inteligentny mężczyzna*, and is represented by 'a box' with predicates inside.

The predicate exhausts the meaning of NP in the case of *predicative uses* e.g.
(1) *Jan jest przystojnym, inteligentnym mężczyzną.* [*Jan is a handsome, intelligent man*]
However, in most cases, NP introduces information about some objects into the domain of discourse. Registration of the introduced objects is done by the *referent activation operator*. In the discourse representation, it is always used in the *SLS* expression, called further *activating expression*, binding the value of the activated referent with some noun predicate. Graphically this expression, of type ((*et*) (*s*(*st*)) is represented on Fig. 3.
The dotted line converts the graphical expression into some kind of lambda expression, where the argument (the lacking element to be delivered) is defined by its shape[3]. In Fig. 3, a noun predicate has to be delivered (double line parenthesis means the meaning of the expression).
During generation of the output state, the activating expression adds (*activates*) to the input state a new referent by changing the most recently activated referent of the output state on the successor of the one from the input state. From the definition of state, the activated referent is accessible. The set of accessible referents is affected by dynamic negation and implication in a standard way for dynamic semantics e.g. Groenendijk & Stokhof (1991).
The referent activation operator assigns indeterministically to the just activated referent some set of objects as a value. The value of the activated referent is to be constrained by the predicate being the argument of the lambda expression. The expressions describes a relation on states, where for the given input state we can have many output states differing in the value assigned to the just activated referent[4]. Additional *compartment* of the 'box' can be used for the name of the referent facilitating comprehensibility of the representation. However, the names of referent play no role in *SLS*. The only tool binding sub-expressions are links.



e.g. ||*jakiś*|| in ***jakiś*** *mężczyzna*

Fig. 3 Expression of the referent activation operator.

In order to not introduce two different kinds of referents: one for singular NPs and one for plural, each referent is always assigned a set f objects – in case of referent introduced by singular NP the set always consists of only one object.
Representation of anaphoric NPs, both definite and pronouns, besides the referent activation needs to include an operation of creation of the link between a given referent and its antecedent. The operation is performed by the *reference operator* which looks for 'appropriate' referents among accessible in the input state and adds a link for each of them to the set of pairs of the output state. The name of the operator originates from the definition of reference adopted in Piasecki (2002b) and in slightly modified version stated earlier. The reference operator is constrained by two necessary conditions:

- accessibility relation expressing syntactic-semantic structure of discourse,
- and 'appropriateness' of the potential antecedent, which is based on some kind of *subsumption* (intended generalisation): the value assigned to the potential antecedent must be included in the denotation of the predicative expression given to the reference operator as its argument.

Reference operator is associated with one of the conditions of existential presupposition:

- *strict* (typical for demonstratives), where the presupposition modifier expects that exactly one link has been added to the links of the output,

[3] The lamba operator together with a name of an argument can be used overlty in the case of confusion.
[4] In that way, assignment have a character of existential import.

- and *weak* (typical for *referential indefinites*, e.g. Kamp & Reyle (1993) or *crypto-indefinite* of Grzegorczykowa (1995), where any non-empty change suffices.

In second case, the speaker can identify an appropriate person, but he does not expect the hearer to be able to do the same. Anyway, the hearer is expected 'to try' to do it. He can create many potential links, transforming the given input state into many possible output states. This property is captured in *SLS* by the axiom of *Indeterministic Interpretation of Reference* mentioned earlier, i.e. in each state the value assigned to a referent must be equal to at least one of the referents 'pointed to' by the referential links

If we assume a non-empty initial context, which can be described by *SLS* expressions as well, the same reference operator can be used in the interpretation of referential NP.

In discourse representation the reference operator is always associated with the presupposition modifier and combined with activation expression forming and *SLS* expressions of type ((*et*) (*s*(*st*)) presented graphically on Fig. 4.

e.g. ||*ten*|| in ***ten*** *mężczyzna* e.g. ||*pewien*|| in ***pewien*** *mężczyzna*

Fig. 4 Expressions of the referent activation operator.

On Fig. 4, the double line arrow signals the presence of the reference operator and the numbers ascribed to it express the conditions introduced by the presupposition modifiers. When the given referential or anaphoric NP is analysed in the context of the whole discourse, the referential arrows can be connected with referents or antecedents, respectively. However, this is not necessary and has only meaning for the reader. Reference operator works autonomously. Assignment of sets of objects (instead of individuals) to referents is the base for unified treatment of (semantically) plural and singular NPs. Singletons represent individuals. Following *CNDM* system of van der Does (1994) and classical solution of Barwise & Cooper (1981), quantifying lexeme is not a generalised quantifier (GQ) but a functor of type ((*et*) ((*et*) *t*)), called further *proto-quantifier*, producing GQ. For example, on the first diagram from the left in Fig. 5, a proto-quantifier **three** takes a value of the referent as an argument and produces a family of all possible subsets of the given set, where each subset contains three elements.

Proto-quantifiers forming relative quantifiers, e.g. **most**, take the denotation of the given noun predicate as their argument e.g. ***większość*** *mężczyzn*. Relative quantifiers introduce a presupposition of non-empty base set from which they generate subsets, e.g. Kamp & Reyle (1989).

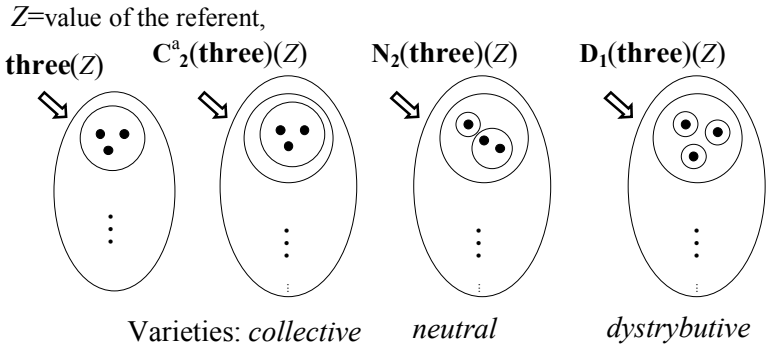


Fig. 5 Diagrams illustrating the work of proto-quantifiers and the variety modifiers

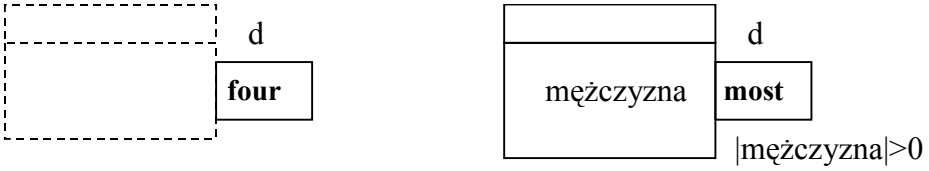
However, it is not the whole truth about the quantification in natural language. We must fur-

ther distinguish between three different *varieties* of quantification, see van der Does (1994), Polish examples come from Grzegorczykowa (1995):

- *distributive* – where the GQ describes the size of set of objects separately involved in some action e.g. (2) ***Każde*** *dziecko uśmiechnęło się.* [***Each*** *child smiled*]
- *collective* – where the GQ describes size of a *collection* of objects performing collectively some action (3) ***Wszyscy*** *mieszkańcy zgromadzili się na placu.* [***All*** *citizens gathered on the square.*]
- *neutral* – where the GQ describes size of a set of objects performing some action in 'sub-groups' e.g. (4) *Pięciu mężczyzn przeniosło dwie szafy.* [Five men moved two]

Technically, collection is defined by van der Does (1994) as a set of objects fulfilling some noun predicate. Variety is modelled with the help of *type lifts operators* of type (((*et*) ((*et*) *t*)) ((*et*) (((*et*) *t*) *t*))), which transforms proto-quantifiers to functors taking a set and returning a family of sets of collections. Collections consisting of only one object are called atomic and represent individual entities. There are appriopriate formal modifiers for each value of variety. Their work is illustrated on Fig. 5.

Proto-quantifiers are graphically represented in the form of additional 'box' (see Fig. 6) with one of the letters: *d*, *c*, *n*, describing the value of variety.



e.g. ||*cztery*|| in ***czterech*** *mężczyzn* e.g. ||*większość*|| in ***większość mężczyzn***

Fig. 6 The representation of absolute proto-quantifier (left) and NP containing relative proto-quantifier (right).

In the case of relative proto-quantifiers, the presupposition condition of non-empty set is written below the box (on Fig. 6 the meaning of whole NP is presented in order not to blur the diagram with the details of 'graphical lambda' expression).

In *DRT* the semantics of a sentence is modelled as following:

- we ascribe referents to NP, putting descriptive constraints on them,
- next, the same referents are use in a predicate representing the meaning of VP,
- finally, these uses of referents are joined together by means of quantifiers and their scope.

Because there are no variables on referents in *SLS*, the picture is very different there:

- proto-quantifiers present in each NP 'produce' family of sets of collections,
- *cardinality dependence relations* between subsequent pairs of quantifiers (full NP) define a set of possible structures of relation, called *collections configurations*,
- and finally on the level of the meaning of a simple sentence, a test is performed whether at least one collection configuration is fully included in the set of collections configurations being the denotation of the verb predicate[5].

The notion of cardinality dependence originates from the work of Bellert (1989) on graph representation of quantified sentences (generally limited to the distributive use of quantifiers) and an attempt to its formalisation done by Zawadowski (1989). Having in mind that the meaning of NP is represented as a set of sets of collections, we can combine representations of subsequent NPs typically in several different ways, e.g. the sentence with both NPs used distribu-

[5] The set can be understood as describing all possible situation for whose description the verb can be used.

tivelly, describing a relation involving individuals (represented as atomic collections):
(5) ***Trzech*** *profesorów sprawdziło* ***dwie*** *prace.* [***Three*** *professors checked* ***two*** *papers*]
can describe the types of situation shown on Fig. 7, on left and right side of the vertical line.
In Fig. 7 dots represent objects, circles and ovals represent sets. On the left part of Fig. 7, we can see the interpretation in which NP(*dwie prace*) depends cardinally on NP(*trzech profesorów*), probably the most preferable reading, on the right, an opposite interpretation is presented. Two NPs can be also cardinally independent. In that case, there are exactly two sets of collections in each configuration collection, where their members are connected in some way. The dependence is not limited to the distributive varieties, but not all possible combinations of directions of dependencies and values of variety are supported by language data e.g. any NP can not be depended on NP used in neutral variety. Moreover, the structure of dependence relation in case of more than two NPs must preserve some general rules expressed in *SLS* in the form of axioms originating from the informal rules proposed by Bellert (1989).

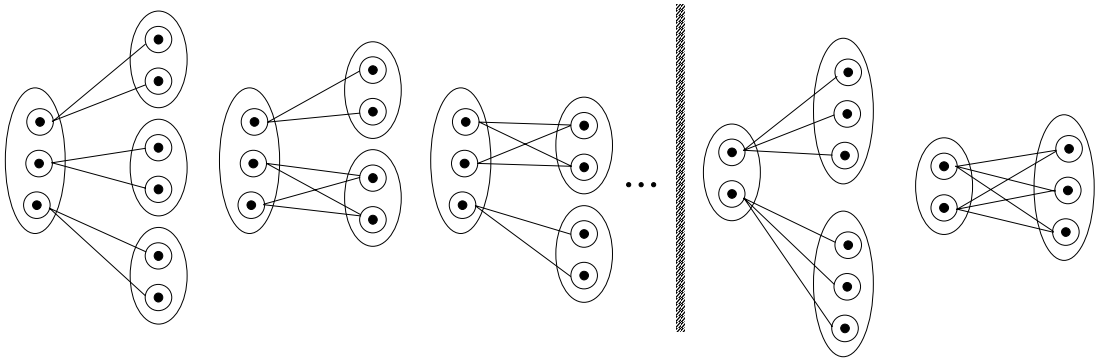


Fig. 7 Diagram illustrating possible structures of relation for distributive interpretation of both NPs in (5) corresponding to: 'wide' scope of *trzech profesorów* on the left and 'narrow' scope on the right.

Cardinality dependence and independence is expressed in *SLS* with the help of binary operators, which applied to quantifiers produce all possible collections configurations in the given class of graphs e.g. the left part of Fig. 7. In the case there are more than two NPs in one VP (a flat structure of Polish VP is being assumed here, see Przepiórkowski et al. (2001)), the dependence relations are stated in *SLS* binary for each pair of NPs. In that way the dependence operators constrain partially the final set of all possible collections configurations for the given VP. The final set of configurations is generated by the *matrix operator* of the appropriate valence, taking as its arguments a sequence of dependence operators and a sequence of products of proto-quantifiers and returns a set of possible collections configurations of type $((et)^{i}\ t)\ t)$, where $i \in \mathbf{Nat}$.

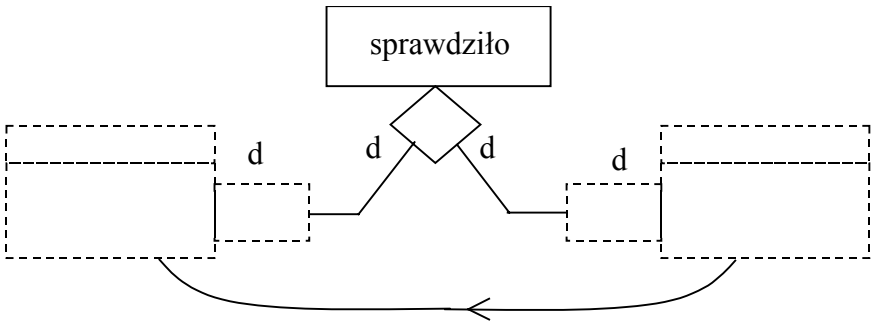


Fig. 8 Interpretation of verb as comparison of possible collections configurations generated by quantifiers with denotation of verb predicate.

In order to deliver to a matrix operator the quantifier, meaning of NP is represented as pair: an expression of type $(s(st))$, being an activation expression or reference operator expression (de-

scribing the connections of the NP with the context) and a quantifier – a product of modified proto-quantifier of type (((*et*) *t*) *t*).
Denotation of verb predicate is also of type ((*et*)^{i} *t*) *t*), where *i* equals to its valence, but this time it describes 'real' situations: all collections configurations for description of which we can use the given verb. Compatibility of the structure, described by quantifiers and their dependence, with the denotation of the verb is tested by *intersecting operator*: their 'common' part have to contain at least one collection configurations consisting of exactly (and only) of objects assigned to the referents introduced by the given NPs.
The meaning of the verb is represented graphically on the Fig. 8 (above). On Fig. 8, diamond means the sequence of operations: combination of quantifier part of meanings of NPs into the set of collections configurations done according to dependence operators, and next, comparison of the result with all 'possible situations' expressed by the denotation of the verb predicate[6]. The single arrow expresses the cardinality dependence relation and its direction. In the case of independence no link is written.
For sentence (5), combining the meaning of the verb with the meaning of its arguments we get the *SLS* expression graphically represented on Fig. 9.

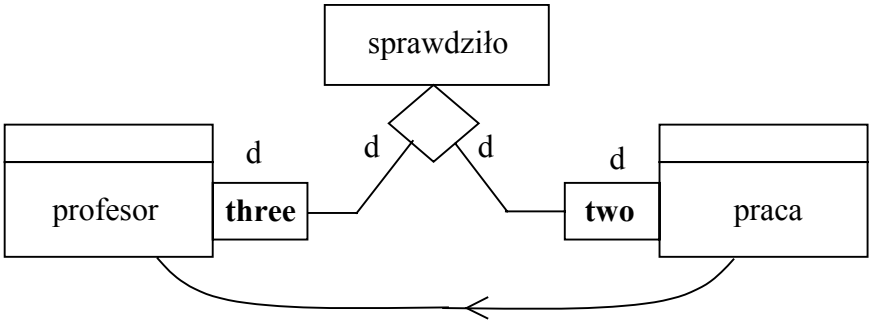


Fig. 9 Representation of one of the possible meanings of (5) i.e. the meaning with 'wide scope' interpretation of NP(*czterech profesorów*).

Fig. 9 presents interpretation, in which cardinality dependence relations are representing 'the wide scope reading of NP(*czterech profesorów*) i.e. there are four professors and each of them separately checked two papers. Dependency structure is introduced by verb meaning in an underspecified way: the verb generates the full set of all possible values of dependence operators preserving the axioms of *SLS* and agreed values of variety.

4. Overview of The Existing Approaches to Polish NP Syntactic-semantic Structure

There is very few formal analysis of the meaning of Polish NP in the literature. The broad base of approaches concerning a syntactic structure and meaning of Polish NP can be dived loosely into four main groups:

1. structural-functional analysis - Topolińska (1984), Grzegorczykowa (1996), concerning generally syntactic issues but bringing a lot of semantic observations,
2. purely semantic, partially formal, e.g. Grzegorczykowa(1995), Kosseska-Toszewa (1992),
3. informal or semi-formal syntactic analysis e.g. Saloni & Świdziński (1998),
4. and, formal, syntactic descriptions done in:
   - *DCG* formalism of surface grammar: Szpakowicz (1976), Świdziński (1992),
   - *Government and Binding Theory*: Węgrzynek (1995),
   - *HPSG*: e.g. Marciniak (2001), Przepiórkowski (1998), Przepiórkowski et al (2001).

[6] The exact form of the verb has been left, because the present treatment of the VP meaning is simplified in *SLS* and will ammended in the future.

In addition to the above classification, some interesting observations originating from cognitive experiments and concerning inter-relations between referentiality and word order are given by Szwedek (1976).
From the analysis in 2), the following classification (slightly simplified here) of NP uses emerges (in parenthesis the names introduced by Grzegorczykowa (1995) are given):

1. predicative (non-referential/predicative): in predicative sentences with *być* e.g. (1)
2. argument/'designating' Polish: *wyznaczające* (referential)
    - 2.1 non-objective (*no such category*), e.g. denoting masses, substances
    - 2.2 objective ('specific' Polish *szczegółowe*)
        - 2.2.1 identifying (definite): PN, demonstratives, definite descriptions
        - 2.2.2 non-identifying (indefinite) overtly subcategorised by Grzegorczykowa into: indefinite for both, crypto-indefinite (Topolińska: "modally marked"), indefinite limited to some set, indefinite element of some class[7]

Moreover, Grzegorczykowa, according to her treatment of quantification as a kind of reference, introduces "general" as an additional subcategory, whereas Topolińska perceives quantity (i.e. quantification) and plurality ('indefinite information about cardinality'), and as parallel features to the above classification. However, in both approaches we can find distinctions between distributive and collective uses, both notice different forms of expressing cardinality and both identifies generic uses as a different form of speaking about general information.
All the distinctions of the classification can be identified with values of different features recapitulated in section 2 and next formalised with *SLS*, e.g. 1 vs. 2 equals generic vs. specific, 2.2.1 vs. 2.2.2 equals referential vs. attributive, different forms of definiteness corresponds to different forms of semantic subsumption and different presupposition conditions.
Besides constructing the detailed classification, Topolińska outlines possible shape of NP syntactic structure, unfortunately appealing mainly to semantic motivations. She divides NP into:

- RQ=[lexical markers of reference and quantity judgement] + KP=[kernel phrase], where:
    - RQ → RF=[reference marker (*operator*)] + QT=[quantity judgement markers (*quantifying word*)] +[unit name],
    - KP → [predicative and argumentative expressions used as attributive expressions] + CC=[constitutive constituent].

Most of the elements but one – CC – are optional. CC can be PN, PRON, NOUN or, in some case, ADJ. QT encompasses also many types of numeral including "improper" like *kilka*, *trochę, wiele*. Topolińska also notices that an important 'constituent' of QT is "morphological category of number". RF includes mainly *demonstrative pronouns* but also some quantifying lexems, e.g. *każdy*, *wszyscy* (differently classified in literature). RF and QT typical (unmarked cases) can appear on the initial position with mostly fixed linear relation between them.
Grzegorczykowa (1996) follows mainly observations in the area of NP of Topolińska while introducing very little information to the syntactic structure of NP. Existing *DCG* descriptions also reveals very little about the position of lexemes of 'type RQ' in the structure of NP. Many of them are excluded from the subset of Polish being described.
Some interesting suggestions deliver formal papers of Węgrzynek (1995) and from 'HPSG area', especially Przepiórkowski (1998), which are not primarily oriented on NP structure.
Węgrzynek starts with a set of syntactic criteria as a base for distinguishing lexemes belonging

[7] This subcategory includes such pronouns like : *jakikolwiek*, *którykolwiek* etc. Grzegorczykowa defines their use as « referening to an indefinite element of some definite class » (but not set !). However, it can be noticed that they can be used only in 'intentional' contexts like : questions, orders, future, arguments of intentional verbs. Their role seem to be emphasasing the 'imaginery' character of the represented object.

to the category of adjectives. As one of the results, she identifies a group of lexemes, morphologically: adjectives or adjectival pronouns, which modifies noun, takes a initial position in NP but do not fulfil the syntactic criteria. She proposes to form a new syntactic category called *determiners* including:

(6) demonstrative pronouns like: *ten* [*this*], *tamten* [*that*], *ów* [*that*] and "adjectival pronouns" like: *jakiś* [≈*a*/*some*], *pewien* [≈*a*/(stressed)*some*], *żaden* [*any*], *każdy* [*each*], *wszyscy* [*all*], *niektórzy* [≈*some*].

Węgrzynek also introduces a determiner phrase (DetP) into NP structure.

The case agreement properties of numerals cause, that in most of the recent approaches, instead of NP including numeral as adjunct, we have numeral-nominal phrase (further NumNP) with numeral phrase as a head and nominal argument e.g. Przepiórkowski (1998), Saloni & Świdziński (1998) and Derwojedowa et al. (2002). In the last approach, besides syntactic analysis of numeral phrases as "sequences of morphological numerals", "an irregular adjectival phrase", corresponding to DetP of Węgrzynek, is signalled on initial position of NumNP.

Przepiórkowski (1998) analysing syntactic properties of numerals and numeral-like lexemes argues that, on the base of strict syntactic criteria, numerals can be subdivided into 3 subcategories of numerals: *pięć*-type (*five*-type), paucal, collective and numeral-like into: numeralisations and indefinite numerals, being "those lexemes which denote a certain imprecise quantity" and which pattern with other numerals in syntactic behaviour. Indefinite numerals, what is important for the rest of the paper, can be further subdivided into four "classes": C1 "which parallel with *pięć*-type" (e.g. *kilka* [*several*], *parę* [*a couple*], *wiele* [*many*]), C2 collective (e.g. *kilkoro* [≈*several*], *kilkudziesięcioro* [≈*several*]), C3 "which behave like *tysiąc* [*thousand*]" (e.g. *szereg* [*array*], *mnówstwo* [*lots of*], *większość* [*majority*]) and C4 "with restricted paradigm" (e.g. *dużo* [*a lot*], *mało* [*little*], *trochę* [*a little*]).

5. Syntactic-semantic Structure of Polish NP

According to the semantic representation of a simple sentence (Fig. 9) there are two mechanisms, which should be expressed in the structure of every nominal phrase: referent activation and some form of quantification. Looking for their lexical sources, we will start with the short inspection of syntactic structure, based initially on proposals adopted from Przepiórkowski and Derwojedowa et al. and extended with DetP category of Węgrzynek:

1) NmP → DetP NumNP

The label NmP means 'Nominal Phrase' and originates from the idea of Świdziński & Saloni (1998) of "nominal phrase" realised as "groups of many types", among them "numeral group", a notion similar to NumNP. The structure in 1) has a temporary character and will be amended later in the paper. According to Węgrzynek (rule 2) and Derwojedowa et al. (rule 3) we have:

2) DetP → (ModP) Det 3) NumNP → NumP NKer

In the rule 2), the parenthesis express optional element, ModP is phrase modifying quantity expressions and it can include adverbs and prepositions, e.g.: ***niemal*** *trzy* [***almost*** *three*], ***tylko*** *cztery* [***only*** *four*], ***mniej niż*** *pięć* [***less than*** *five*]. The idea of acceptance of at most one Det is supported by examples 7a to 7c, below:

(7a) **każdy ten mężczyzna* (7b) **ci niektórzy mężczyzni* (7c) ??*ten pewien mężczyzna*
**each this man* **these some men* **this SOME man*

(7d) *wszyscy ci mężczyźni* (7e) *każdy z tych mężczyzn* (7f) *mężczyzna*/*mężczyźni*
**all these men* *each of these men* **man* (without article) */ men*

The case (7d) seems to be a counterexample but, as it will be argued later, (7d) is due to its semantic structure similar to (7e). The difference between them lays in the value of variety.

In the rule 3), the original label 'NP' has been exchanged here, in order to avoid confusion.

The label for this part (i.e. NKer) of 'numeral headed nominal phrase' has been given in analogy to proposal of Topolińska. However, as the examples below reveal, not all Dets can be connected to NumNP including NumP:

(8a) *ten mężczyzna* (8b) *tych trzech mężczyzn* (8c) *jakiś pięciu mężczyzn*
this man *these three men* *some five men*
(8d) ??*ten jeden mężczyzna* (8e) *każdy mężczyzna* (8f) **każdych trzech mężczyzn*
this one man *each man* **each three men*
(8g) *dwóch mężczyzn* (8h) ??*wszyscy trzej mężczyźni* (8i) *jakiś mężczyzna*
two men *all three men* *a man*

According to the mechanisms of *SLS*, each specific (i.e. non-generic) NmP can introduce only one discourse referent. Because there can be at most one Det, its meaning representation is an obvious candidate for the referent activation expression of *SLS* to be located in it.

An obvious place, to ascribe quantification mechanism to, is meaning representation of NumP. However, traditionally, Det *jakiś* is also perceived as a typical marker of existential quantification. But, if there is no NumP, then the aspect of quantification should be associated with morphological number of the noun (at least there is distinction between singular object and some unspecified 'number of' objects). Thus in both constructions, there would be two quantifiers in the same NumNP: one in Det and one in NumP. From the technical point of view there are no objections against more the one quantifier and an additional operation of composition of quantifiers could be introduced into *SLS*. However, (8e) and (8h) show that overtly quantifying Dets do not co-occur with NumP.

According to these observations, we can divide determiners into two semantic subcategories: quantifying and non-quantifying. The semantic representation of non-quantifying is presented above: *jakiś* on Fig. 3, *ten* and *pewien* on Fig. 4. All of them activate a new referent, *ten* and *pewien* additionally introduce the reference operator. The difference in meaning between *ten* and *pewien* is described as the difference in presupposition. In the case of *ten*, the hearer is obliged to identify exactly one referent, while in the case of *pewien*, the hearer should try to identify the referent but the speaker do not expect that he would be able to do this precisely. The hearer can have many different 'hypothesis' about the identity of the referent.

Quantifying determiners are 'self-contained' in respect to activation and quantification. They lacks only descriptive content. Moreover, most of them include relative quantifiers, what is suggested by the examples in (9), in comparison to examples in (10):

(9) **Istnieje*(*ą*) ***każdy*** *uczeń/****wszyscy*** *uczniowie w tej szkole, który/którzy...*
**There is(are) each schoolboy/all schoolboys in the school, who...*

(10) **Istnieją* ***wszelkie*** *genialne pomysły na zarabianie pieniędzy.*
**There are* (≈)*all splendid ideas on earning money.*

(11) *Istnieje dwóch uczniów/kilku uczniów/dużo uczniów/jeden uczeń/****niejeden*** *uczeń w tej szkole, który/który...*
There are two/several/a lot of/one/not a few schoolboys, who ..., in the school.

The possible explanation of non-acceptability of (9) can be that relative quantifiers presuppose existence of non-empty base set from which they 'take' a part. The meaning representation of quantifying Dets is presented on Fig. 10. Symbol '{...}' means underspecified value of variety, i.e. the value can be one of the listed in curly parentheses.

Quantifier $\mathbf{exist_{pl}}$ used in Fig. 10, returns a family of all subsets of the given set (value of the activated referent), which contain at least two elements. Quantifier **every** returns a family consisting of exactly one subset identical with the given set (equal to the denotation of noun).

The task of NumP is to introduce a quantifier and, as in traditional approaches, to set the stage preparing positions for all other elements, see Fig. 11.

In Fig. 11, quantifier **three** returns a family of subsets of the given set (value of the referent),

which contains three elements and quantifier **most** returns a family of subsets, which contain more than half of the objects of the given noun denotation[8].

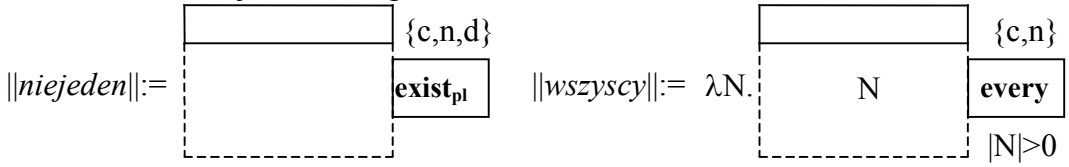


Fig. 10 The semantic representation of quantifying determiners.

Fig. 12 presents compositional construction of semantic representation of 'full' NmNP (8b).

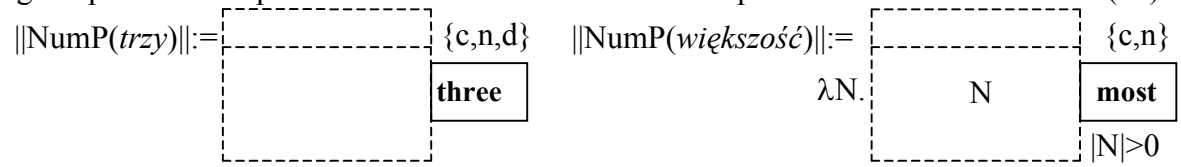


Fig. 11 The semantic representation of numeral phrases.

Examples (8a), (8g), (8i) show that one of the positions DetP and NumP or very often both of them, as in (7f), can be left empty. However, NmP without DetP can be anaphoric antecedent in a simple not negated sentence, like in (12) and (13) and should activate a referent.

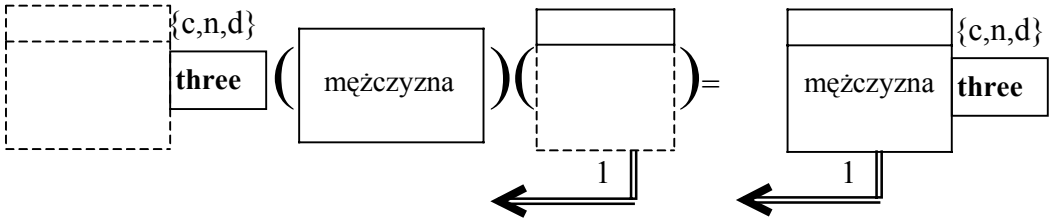


Fig. 12 Compositional construction of nominal phrase (8b).

(12) *Na rogu ulicy stoi* ***budynek***[1]*. Pomalowano* ***go***$_1$ *w żółto-brązowe pasy.*
There is ***a building***[1] *on the corner.* ***It***$_1$ *is painted in yellow-brown strips.*

(13) ***Dwóch profesorów***[1] *sprawdziło prace. Nie zdążyli* ***oni***$_1$ *jeszcze ogłosić wyników.* [***Two professors***[1] *have checked the papers.* ***They***$_1$ *have not yet managed to announce the results.*]

NmPs without DetP are very often ambiguous according to its value of referentiality: they can be used anaphorically, e.g. Topolińska (1984), or referentially see Szwedek (1976). But, it seems that NmPs including only NumP, e.g. (14) and (15), can not be referential.

(14) ****Dwóch profesorów***[1] *sprawdziło prace.* ***Dwóch profesorów***$_1$ *opuściło pokój.*
****Two professors***[1] *checked the papers.* ***Two professors***$_1$ *left the room.*

(15) ***Dwóch profesorów***[1] *sprawdziło prace.* ***Tych dwóch profesorów***$_1$***/profesorowie***$_1$ *...*
****Two professors***[1] *checked the papers.* ***These two professors***$_1$***/professors***$_1$ *...*

The lack of DetP can be solved in two ways by introduction of: an ambiguous empty determiner or a *type lifting* operation associated with the appropriate syntactic rule. The second solution, better from the theoretical point of view is applied here. However, because in general case the lack of DetP can be ambiguous: attributive or referential with weak and strict presupposition, we need three different type lifts operators, called: *AL*, *RWL*, *RSL*, changing the type of *SLS* expression from corresponding to the representation of NumP to corresponding to NmP. The semantic definitions of the operators are similar with application of the meaning of *jakiś*, *pewien*, *ten*, respectively, (i.e. *AL*, *RWL*, *RSL*) to the meaning of NumNP (see Fig. 14).
Similar solution is applied to the problem of the lack of NumP in NmP. There are two type lifts operators, called *q-lifts*, transforming the type from type of NKern representation to the type of NumNP representations. However, there is no problem of ambiguity: the number cate-

[8] The sets generated by relative quantifiere are later connected to the value of the activated referent by application of matrix and intersection operators.

gory of NKern differentiates between QL_{SG} and QL_{PL} operator, see their definitions on Fig. 13.

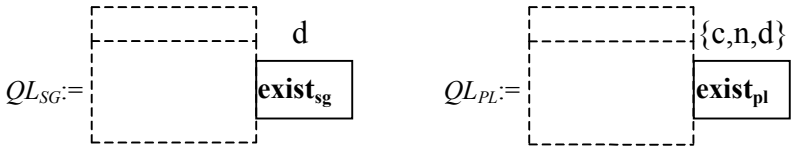


Fig. 13 Definitions of QL_{SG} and QL_{PL} type lifts operators.

In Fig. 13, the definition of quantifier **exist$_{pl}$** is identical to its use in Fig. 10. The quantifier **exist$_{sg}$** returns a family of atomic subsets of the given set (the value of the activated referent). Fig. 14 presents construction of representation of NmP(*mężczyzna*), where both types of type lifting operators must be applied to get the right semantic representation.

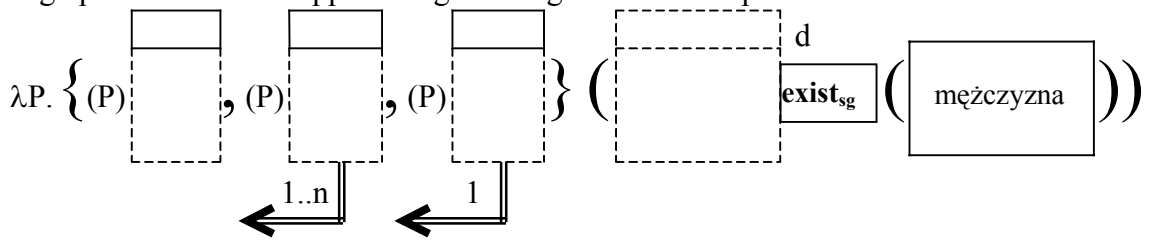


Fig. 14 Construction of underspecified semantic representation of NmP(*mężczyzna*).

According to the semantic structures of different NmP proposed above, the initial syntactic structure of NmP can be amended now in the form of the following simplified rules (where square brackets means list of pairs: attribute-value (+/- means binary values):

4) NmP → DetP[quant-] NumNP[qlifted-]
5) NmP → DetP[quant-] NumNP[qlifted+]
6) DetP[quant] → (ModP) Det[quant]
7) NmP → NumNP[qlifted+]
8) NmP → DetP[qaunt+, num=*x*] NKern[num=*x*]
9) NumNP[qlifted+, num=sg] → NKern[num=sg]
10) NumNP[qlifted+, num=pl] → NKern[num=pl]
11) NumNP[qlifted-] → NumP NKern

Value of *quant* divides determiners in the lexicon into two subcategories: non-quantifying (*quant*-) and quantifying (*quant*-) i.e.:
Det[quant-]:= *jakiś*, *ów*, *pewien*, *ten*, *tamten,* and *inny* [*another*], *taki*,
Det[quant+]:= *niejeden*, *każdy*, *wszelki*, *wszyscy*, *żaden* etc.
The direct creation of NumNP from NKern (rules 8 and 9) is associated with the application of the appropriate *q-lift* operator (depending on the number) in the construction of semantic representation. In a similar way, direct movement from NumNP to NmP (rule 7) involves type lift operator, underspecified (this time) among operators: *AL*, *RWL*, *RSL* (Fig. 14). Quantifying determiners, combining referent activation and quantification in their semantic representation (Fig. 10), join (rule 11) NKern, semantically represented by compound predicate, to form NmNP. Quantifying determiners can not join NumNP because NumNP always include some form of quantification in their semantic representation: quantifier introduced by numerals (rule 4) or created by *q-lift* operator (rule 5).
The rules 4-11 and semantic representation associated to them describe all examples of simple NmP presented so far. But, there are two other classes of simple NmNP, which are not covered by them: including indefinite numerals of class C4, Przepiórkowski (1999) and constructions with prepposition *z* (here: *of*). C4 numerals can not co-occur with determiners, e.g.:

(16a) **tych mało mężczyzn* (16b) **jakiś trzech mężczyzn* (16c) **każdych dużo mężczyzn*
**these little men* **some three men* **each a lot men*

(17) ***Mało studentów***[1] *zdało egzamin. Mimo to,* ***oni***$_1$ *studiują nadal.*
****Little students***[1] *have passed an exam. Anyway,* ***they***$_1$ *continue studying.*

(18a) *mało tych mężczyzn* (18b) **mało jakiś mężczyzn* (18c)[9] **mało pewnych mężczyzn*
**little these men* **little some men* **little SOME men*
(19a) *wszyscy ci mężczyźni* (19b) **wszyscy jacyś mężczyzn* (19c) **wszyscy pewni mężczyźni*
all these men **all some men* **all SOME men*

However, C4 numerals behave in a similar way to Det[quant+]: they create a referent, e.g. (17), and express identical pattern to *wszyscy* in (19a) type constructions (discussed below).
Avoiding problematic solution of the yet another application of *q-lift*, we propose to ascribe to C4 numerals a similar syntactic structure and semantic representation as to Det[quant+] i.e.:

12) NmP[num=pl] → NumC4P NKern[num=pl]

Semantic representation of C4 Num is identical as in Fig 10, with the appropriate quantifier.
The ability of determiner *wszyscy* to co-occur with determiner *ci* like in (19a), distinguishes it from other Dets. But, (19b) and (19c) suggest that only referential NmPs are accepted on these positions. Referential NmPs are also expected in complementary *z*-constructions e.g.:

(20a) *każdy* z *tych mężczyzn* (20b) **każdy z jakiś mężczyzn* (20c) *trzech z mężczyzn*
each of these men **each of some men* **three of the men*

In (20a) and (20c) quantification is performed on the set included in some other set pointed out. In (19a), we speak about the identified set as the whole (improper inclusion).
In both cases, when the expression are anaphoric antecedents, they represent the distinguished part - not the identified whole. These two cases differs only in the value of variety: *z* enforces distributive reading (at least on the left argument), while *wszyscy* is claimed, e.g. Grzegorczykowa (1995), to have only collective reading. Concluding, we need two discourse referents, where one has almost empty descriptive content, see Fig. 15.

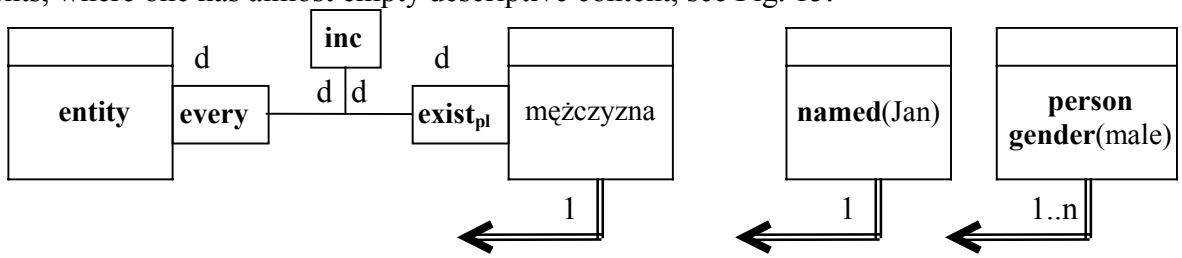


Fig. 15 Representation of *z*-type constructions (20a), proper name *Jan* and pronoun *on* [*he*].

In Fig. 15, **inc** is the relation of type ((((*et*)*t*)*t*) ((((*et*)*t*)*t*) *t*)) accepting two families of sets of collections, where all collections are atomic, the 'right' family contains exactly one set and each set belonging to the 'left' family is included in the single set of the 'right' family. (19a) can be also represented with the help of **inc**, but with the value of variety of the 'left' argument set to *c* (collective). Fig. 15 presents also semantic representation of 'self-contained' lexemes of proper names and pronouns. In the case of pronouns, weak presupposition is assumed, because of the acceptability of discourses with ambiguous anaphoric links.
To complete the picture, it is worth to notice the existence of various modifiers of quantification in the NmP structure e.g. *niemal* [almost], *przynajmniej* [at least], *około* [*about*]. They can be represented as operators changing one proto-quantifier into another one and they can be located in DetP[quant+] and NumP on modifying positions.
Another group of potential nominal variety modifiers, e.g. *razem* [≈*together*], *łącznie* [≈*together*], preposition *po*, seem to be rather variety modifiers of verb than NmP. Most of them can appear only once in a sentence, they can not be attached to subsequent NmPs.

[9] Here, the intended reading of *pewny* is crypto-indefinite marker. If the reading is *reliable* then the sentece is perfectly correct. The same is true for (19c)

6. Conclusion and Future Works

The presented analysis enables compositional description of NP meaning in large extent, but does not cover all its aspects. Besides the detailed analysis of lexical semantics of nouns, there is still left an important area of mass nouns where we need to regard a unit name and to change the interpretation of the *collection*. In the present shape of SLS, collections are just sets. Another interesting problem is the representation of an anaphor to descriptive contents of NmP, e.g. by determiner *taki* (Eng. *such*). In order to do this, we need to extend the linking mechanism of *SLS* with links to predicates (in that way predicates should be represented in 'memory'). Moreover, the semantics of VP is only touched now. The usage of the notion "situation" is not accidental and signals the possible directions of development.

References

Barwise J., Cooper R., Generalized Quantifiers and Natural Language. Linguistics and Philosophy 4:159-219, 1981.

Bellert I. Feature System for Quantification Structures in Natural Language. Foris Pub., 1989.

van der Does J. Applied Quantifier Logic, Doctoral dissert., ILLC, Univ. of Amsterdam, 1994.

Derwojedowa Magdalena, Rudolf Michał, Świdziński Marek. Two Formal Approaches to Polish Numeral Phrases Implemented. In preparation, UW, 2002.

van Eijck J., Kamp H. *Representing Discourse in Context*. w van Benthem J., ter Meulen A., editors, Handbook of Logic and Language, Elsevier, 1997.

Groenendijk J., Stokhof M. Dynamic Predicate Logic. In Linguistics and Philosophy 14, 1991.

Grzegorczykowa R. Wprowadzenie do semantyki jezykoznawczej. PWN, 1995.

Grzegorczykowa R. Wykłady z polskiej składni. PWN, Warszawa, 1996.

Kamp Hans, Reyle Uwe. From Discourse To Logic. Kluwer, 1993.

Koseska-Toszewa V. The Semantic Category of Defniteness/Indefinniteness in Bulgarian and Polish. Slawistyczny Ośrodek Wydawniczy, Warszawa, 1992.

Marciniak M. Algorytmy implementacyjne syntaktycznych reguł koreferencji zaimków dla języka polskiego. PhD dissertation, IPI PAN, 2001.

Piasecki M. Compositional Approach to Nominal Anaphora in Dynamic Setting. Report P-10/02, Computer Science Dep. Wrocław Univ. of Technology, 2002(a).

Piasecki M. A Dynamic Approach to Quantification and Reference in Polish Noun Phrase. In, G. Zybatow et al. eds. Current Issues in Formal Slavic Linguistics, Peter Lang, 2002(b).

Przepiórkowski Adam Case Assigment and the Complement/Adjunct Dichotomy A Non-Configurational Constraint-Based Approach. PhD dissertation, Universitä Tübingen, 1999.

Saloni Z., Świdziński M. Składnia współczesnego języka polskiego. PWN, 1998.

Przepiórkowski A., Kupść A., Marciniak M., Mykowiecka A. Formalny opis języka polskiego – teoria i implementacja. In preparation, IPI PAN, 2001.

Szpakowicz S. Formalny opis składniowy zdań polskich. Wyd. UW, Warszawa, 1983.

Szwedek A.J. Word Order, Sentence Stress and Reference in English and Polish. Linguistic Research Inc., Edmonton, 1976

Świdziński M. Gramatyka formalna języka polskiego. Wyd. UW, Warszawa, 1992.

Topolinska Z. Skladnia grupy imiennej (w:) Gramatyka wspolczesnego jezyka polskiego, t. 1, Skladnia, Warszawa, 1984.

Węgrzynek K. Składnia przymiotnika w ujęciu generatywno-transformacyjnym. PAN IJP, Kraków, 1995.

Zawadowski M. System for Quantification Structures in Natural Language. Foris Pub., 1989.