Russian syntax and semantics

A few decades ago I published several partial descriptions of modern Russian (1972, 1973a, 1973b, 1974, 1980, 1986) against the background of a theoretical framework which may be called radical structuralism and which I never presented explicitly in a coherent way. The basic thought behind this approach is that a sharp distinction must be made between the axiomatic foundation of a framework, the creative liberty allowed within the framework, and the observations which are relevant to possible alternatives within the framework. This distinction has important methodological implications. Statements which are logical corollaries of the axiomatic foundation have the status of God’s own truth within the theoretical framework whereas statements which are subject to the researcher’s freedom of choice have the status of hocus pocus explanations (cf. Joos 1957: 80). While one explanation can be simpler or more economical than another, such a choice between alternatives is impossible in the case of properties which follow logically from the axiomatic foundation. Consequently, observations can only be relevant if there is a choice between alternatives and will be brushed aside when they do not fit into the theoretical framework.

Formal grammar starts from the assumption that people generate formal structures which can be filled with (phonetic or semantic) substance when they meet the environment. As a consequence, observations of (phonetic or semantic) data are relevant only to the extent that they fit into a formal structure, which itself is independent of such observations. Principles and parameters of the formal structure can only be established by a high priest who licenses the performance of his followers and regulates their freedom of choice accordingly. The logical development of such a framework is toward minimalist principles and parameters on the one hand and procedural constraints to achieve optimal consistency on the other. The actual linguistic data are largely irrelevant in this approach because they have no bearing on the formal structure.

In the descriptive framework adopted here, the basic assumption is that linguistic communication is achieved through correlating neural maps reflecting visual and auditory aspects of the outside world (cf. Ebeling 1978: 37, Kortlandt 2003: 242). The correlation between phonetic signals and semantic maps implies the existence of minimal
differences on one level which are correlated with some difference on the other. As the speech flow proceeds in time, successful communication is accomplished by the addition of new images to the world view of the receiver. Since unique signals cannot be interpreted, the correlation must be established by pattern recognition. This in turn requires the existence of units which can be recognized. It follows that there are three levels which are inherent in linguistic communication, viz. the level of speech signals which can be correlated with new images (the phonemic level), the level of images which can be communicated through correlation (the semantic level), and the level of correlated units, i.e. of linguistic signs (the morphemic level). These are God’s own truth levels in the present framework because they follow logically from the view of language as a communicative system. Note that there is no room for considerations of simplicity, economy or pattern congruity here because these presuppose a choice between alternatives which is not allowed in a strict application of the principle that communication is achieved through correlation of neural maps. There is no reason to suppose that correlation proceeds in a simple or economical fashion. In fact, the absence of simplicity and economy can be a major nuisance in the real world.

The description of an actual linguistic system requires four other levels of analysis because the three levels mentioned earlier are neither open to direct observation nor subject to logical investigation. Observation of the phonetic and semantic substance implies the existence of a level where the speech flow is described (the phonetic level) and a level where the outside world is described (the pragmatic level). These levels are arbitrary in the sense that more detailed observation of the data requires a higher level of specificity. There is no natural limit here because it cannot be known in advance which features will be relevant to the phonemic and semantic properties of a linguistic system. The latter can only be approached by means of hypotheses about the correlation between phonetic signals in the speech flow and semantic maps reflecting the outside world. These hypotheses are subject to emendation and rejection in favor of alternatives and therefore belong to a hocus pocus level of explanation through logical investigation of the data. In the framework advocated here, there are two such levels, one for the analysis of phonetic signals which can be correlated with images of the world (the morphonemic level) and one for the analysis of semantic maps which can be correlated with the speech flow (the syntactic level). These are levels where consistency, simplicity, economy and pattern congruity play a major role while the data are simply regarded as given.

The computer synthesis of Russian verb forms in ALGOL 60 which I published 35 years ago (1972) represents a stricter and more detailed generative analysis of the flexional system than any alternative which has
come to my attention. It clearly belongs to the morphonemic level. In order to elucidate the differences between the phonetic, phonemic and morphonemic levels, I published a succinct description of Russian phonology and morphology accompanied by phonetic, phonemic and morphonemic transcriptions of a single text (1973a and 1974, cf. also 1973b and 1986). While I have also published detailed analyses of specific problems in Russian (1980), Japanese (1992) and Chinese (1998) syntax and semantics, I have never publicly discussed the generalities involved (but cf. 1984). The reason for this is that Carl Ebeling’s magnum opus (1978) was going to be followed by an application of his theory to an actual text, but this plan never materialized, evidently because the complications were prohibitive (cf. Ebeling 1984 for an illustration of his methodology and 2006 for a further elaboration of the theory and its application to Dutch data). It appears that his theory, which remains the only elaborate framework geared to God’s own truth semantics in the sense explained above, does not easily lend itself to practical application. It is therefore time to present a less ambitious effort to describe Russian syntax and semantics against the background outlined here.

The main tenet of Ebeling’s theory, to which I subscribe, is that semantic maps consist of projections of (sets of) identifiable features carried by identifiable entities in the real world and of their interrelations. It follows that a semantic map can be viewed as a matrix consisting of columns of (sets of) features and rows representing entities carrying them connected by various relations. The following examples may serve as an illustration (cf. Ebeling 1978: 305 and Kortlandt 1980: 244f.).

(1) *She likes yellow tulips.*

This is the assertion (.) of a situation Σ in the present (−s) where an identifiable female person (*she*) is involved in an event (*like*) with a complementary entity which is a set (−s) of elements (*tulip*) which are limited by an additional quality (*yellow*). In Ebeling’s notation:

\[ Σ / \text{PRES} . \text{ASS} \]
\[ \text{she} = [\text{liking}] \]
\[ [\text{liked}] ; \text{tulip} − \text{yellow} / \text{PL} \]

The same features carried by the same entities but connected through different relations are found in the following:

(2) *She likes tulips yellow.*
Here the limiting quality refers to the object of [liking], which has a temporal dimension, rather than to the complementary entity itself:

\[
\Sigma / \text{PRES} \cdot \text{ASS}
\]
\[
she = \text{[liking] [liked], yellow ; tulip / PL}
\]

with “temporal gradation” (,) replacing “oriented limitation” (−) because the quality of being yellow conditions the event of liking rather than its carrier. The Russian translation of (2) is the following:

(3) *Ona ljubit tjul’pany želtye.*

The analysis of this sentence is the same as that of its English equivalent except for the fact that the ending –ye of želtye is not accounted for. This is important because there is an alternative:

(4) *Ona ljubit tjul’pany želtymi.*

Here the substitution of the instrumental želtymi for the accusative želtye gives the impression that the tulips have been painted. The appropriate analysis of this sentence is the following:

\[
\Sigma / \text{PRES} \cdot \text{ASS}
\]
\[
she = \text{[liking] [liked] ; tulip ~ yellow / PL}
\]

with “temporal limitation” (~) expressing that the tulips being yellow must be contrasted with a situation where they were not yellow. A natural example of this interpretation is the following, referring to trees which change their color according to the seasons:

(5) *Ona ljubit derev’ja želtymi.* “She likes the trees yellow.”

It is clear that Russian offers more possibilities than English here because it has a richer morphology.

A reduction of Ebeling’s system of God’s own truth semantics to a generative system of hocus pocus syntactic rules requires a different formalism than the usual type of generative grammar (cf. Ebeling 1978: 502f., Kortlandt 1984: 184). There are two reasons for this. First, Ebeling’s semantic maps reflect not only meaningful (sets of) features but also meaningful relations between (sets of) features. Second, his (sets of) features are distributed over different carriers. As a result, the usual
bifurcations are replaced by more complex configurations. Consider the following example:

\[ S \rightarrow \text{NP} \ \text{VP} \]
\[ \text{VP} \rightarrow \text{V} \ \text{NP} \]

In Ebeling’s framework, the relations between subject and predicate (“nexus”) and between verb and object (“complementation”) are meaningful themselves, so that these rules must be replaced by rules of the type

\[ C \rightarrow \text{A} \ \text{R} \ \text{B} \]

where the relation R has its own semantic contribution to the meaning C, in addition to the (sets of) features A and B. Moreover, features are split into “valences” when they are distributed over different entities, which requires rules of the type

\[ P \rightarrow [Q_1] \]
\[ [Q_2] ; \ \text{A} \]

where A fills the complementary valence of P. Thus, we arrive at a system which looks as follows:

\[ \Sigma \rightarrow \Sigma \]
\[ \text{SUBJ} = \text{PRED} \]
\[ \Sigma \rightarrow \Sigma / \text{CIRC} \]
\[ = \text{PRED} = \text{PRED} \]
\[ \text{PRED} \rightarrow [V_1] \]
\[ [V_2] ; \ \text{OBJ} \]

and so forth. The complexity of this system is a direct consequence of the requirement that the distribution of (sets of) features over their carriers be reflected in the semantic analysis.

Recognizing the God’s own truth character of the semantic level and seizing the opportunity to adapt the system at will in order to arrive at a manageable description of Russian syntax, I now simplify the system by substituting formal symbols and relations for meaningful elements on the basis of simplicity, economy and pattern congruity in the same way as I substituted morphonemes for phonemic units in my description of the morphology (1974). This involves three operations where semantic
distinctiveness is lost on the syntactic level, just as phonemic
distinctiveness was lost on the morphonemic level. Firstly, the meaning
of the semantic relation R in rules of the type

\[ C \rightarrow A \ R \ B \]

must be distributed over the elements A and B between which the relation
holds. This problem is comparable to the dissolution of joint features in
phonology (cf. Ebeling 1978: 77–79), e.g. in Polish [sf] and [tf], where
the phoneme /v/ is devoiced after /s/ and /t/ in \( \text{śwój} \) ‘one’s own’, \( \text{twój} \)
‘your’ while /z/ and /d/ are devoiced before /f/ in \( \text{sformalizować} \) ‘to
formalize’, \( \text{odformalizować} \) ‘to un-formalize’, but not before /v/,
e.g. in
\( \text{zwójka} \) ‘tortricid’, \( \text{dwójka} \) ‘two’, where voicedness is distinctive twice.
Thus, the relation ‘–’ in

\[ \text{tulip} – \text{yellow} \]

can be split into ‘limited’ characterizing ‘tulip’ and ‘limiting’
characterizing ‘yellow’, and the relation ‘/’ in

\[ \text{tulip} / \text{PL} \]

can be split into ‘belonging to a set’ characterizing ‘tulip’ and ‘being a
set’ characterizing ‘consisting of more than a single member’. Note that
both members of the relation have the same carrier in these instances
because they refer to the same portion of the real world, which carries the
image of “yellow tulips”.

Secondly, the distribution of the (set of) features Q over two carriers
in rules of the type

\[ P \rightarrow [Q_1] \]
\[ [Q_2] ; A \]

can be indicated by numbering and indexing the carriers of features, e.g.
\( Q_{1+2} \) for an element with two valences and \( A_2 \) for the element which fills
the second valence. A slightly different example is the reformulation of

\[ \Sigma \]
\[ = P \]

as \( S_{0+1} \), which denotes the situation that is predicated, and \( P_{1} \), which
denotes that P is the predicate. Thirdly, morphemes often lose (part of)
their meaning in syntactic constructions. This is the counterpart of
neutralization on the phonemic level. When distinctiveness gives way to unification on a hocus pocus (morphonemic, syntactic) level, descriptive categories replace units of form and meaning, e.g. in

(4) Ona ljubit tjul’pany želtymi.
(3_{1}-Nsf_{1} S_{0+1} l’ubi_{1+2}-PRES_{0}-3s_{1} t’ul’pan_{2}-Ap_{2} žolt_{2}-Ip_{2} ASS_{0})

which is now the syntactic representation reflecting the semantic analysis

\[ \Sigma / \text{PRES . ASS} \]
\[
\text{she = [liking]} \\
[\text{[liked]}] ; \text{tulip \sim yellow / PL}
\]

cited above. Here (3-sf) corresponds to ‘she’, (N...S...3s) to ‘\Sigma’ and ‘=’, (l’ubi-) to ‘[liking]’ and ‘[liked]’, (PRES) to ‘/ PRES’, (ASS) to ‘. ASS’, (t’ul’pan-) to ‘tulip’, (žolt-) to ‘yellow’, (A) to ‘;’, (l) to ‘\sim’, and (p...p) to ‘/ PL’. These syntactic categories can have different meanings in other instances, e.g.

(6) On upravljaet mašinoj. “He drives a car.”
(3_{1}-Nsm_{1} S_{0+1} upravl’aj_{1+2}-PRES_{0}-3s_{1} mašin_{2}-Is_{2} ASS_{0})

\[ \Sigma / \text{PRES . ASS} \]
\[
\text{he = [operating]} \\
[\text{[operated]}] ; \text{machine / SG}
\]

where the instrumental case fills a valence without any temporal characterization, so that (I) corresponds to ‘;’ here. Things can easily get more complicated when verbal categories are involved, e.g.

(7) Ona poprosila ego rabotat’. “She asked him to work.”
(3_{1}-Nsf_{1} S_{0+1} poprosi_{1+2+3}-PAST_{0}-sf_{1} 3_{2}-Asm_{2} rabotaj_{2}-INF_{3+2} ASS_{0})

\[ \Sigma / \text{PAST . ASS} \]
\[
\text{she = [asking]} \\
"[\text{[asked]}] ; \text{he} \\
[\text{[asked for]}] ; \Sigma \]
\[\Sigma X = \text{[working]}\]

where the second object of [asking] is a situation where the first object carries the feature ‘working’, so that (INF) corresponds to ‘; \Sigma’ and ‘X =’ here (cf. Ebeling 1984: 104).
Thus, I distinguish seven levels of linguistic analysis which can be exemplified by means of the French word for ‘water’ *eau* [o] as follows:

– on the phonetic level, [o] is an instance of the word in the speech flow,
– on the phonemic level, /o/ is the set of phonetic features capable of distinguishing the word from other words,
– on the morphonemic level, <o> is the description of the form of the word in the speech flow,
– on the morphemic level, {o} is the sign that consists of the form /o/ and the meaning ‘o’,
– on the syntactic level, (o) is the description of the meaning of the word in a syntactic construction,
– on the semantic level, ‘o’ is the set of semantic features which differentiate the word from other words,
– on the pragmatic level, “o” is an object referred to by the word in a situation.

It will be clear that the establishment of correspondence rules between syntax and semantics is a major undertaking and remains an important task for the future.

As an illustration of the syntactic analysis developed here I shall now present a syntactic transcription of the same text that I used in my earlier description of Russian phonology (1973a: 80–82) and morphology (1974: 69f.). In order to simplify matters, I shall leave aspectual, lexical and intonational categories as well as flexion classes and accent classes out of consideration here and use a simplified notation which should be self-evident. Categories: N(ominative), G(enitive), D(ative), A(ccusative), I(nstrumental), L(oactive), s(ingular), p(lural), m(asculine), f(eminine), n(euter), SH(ort adjective), COMP(arative), ADV(erbial), POSS(essive), ET, T(demonstratives), K(interrogative, relative), IND(finite), 1(st), 2(nd), 3(rd person), SE(reflexive), SUCH, WHICH, TIME, PRES(ent), PAST, IMP(erative), INF(initive), GER(und), A(ctive-)P(art)T(iciple), P(assive-)P(art)T(iciple), NE(gation), S(entence).

По причинам, о которых не время теперь говорить подробно, я должен был поступить в лакеи к одному петербургскому чиновнику, по фамилии Орлову. Было ему около тридцати пяти лет, и звали его Георгием Иванычем.

К этому Орлову поступил я ради его отца, известного государственного человека, которого считал я серьезным врагом своего дела. Я рассчитывал, что, живя у сына, по разговорам, которые услышу, и по бумагам и запискам, какие буду находить на столе, я в подробности изучу планы и намерения отца.
Обыкновенно часов в одиннадцать утра в моей лакейской трещал электрический звонок, давая мне знать, что проснулся барин. Когда я с вычищенным платком и сапогами приходил в спальню, Георгий Иваныч сидел неподвижно в постели, не заспанный, а скорее утомленный сном, и глядел в одну точку, не выказывая по поводу своего пробуждения никакого удовольствия. Я помогал ему одеваться, а он неохотно подчинялся мне, молча и не замечая моего присутствия. Потом, с мокрою от умыванья головой и пахнувшей свежими духами, он шел в столовую пить кофе. Он сидел за столом, пил кофе и перелистывал газеты, а я и горничная Поля почтично стояли у двери и смотрели на него. Два взрослых человека должны были с самым серьезным вниманием смотреть, как третий пьет кофе и грызет сахаринки. Это, по всей вероятности, смешно и дико, но я не видел для себя ничего унизительного в том, что приходилось стоять около двери, хотя был таким же дворянином и образованным человеком, как сам Орлов.

У меня тогда начиналась чахотка, а с нею еще кое-что, пожалуй поважнее чахотки. Не знаю, под влиянием ли болезни, или начинавшейся перемене мировоззрения, которой я тогда не замечал, мною изо дня в день овладевала страстная, раздражающая жажда обыкновенной, обывательской жизни. Мне хотелось душевного покоя, здоровья, хорошего воздуха, сытости. Я становился не столько читателем, как читатель, не знал, что, собственно, мне нужно.

(Из “Рассказа неизвестного человека” А. П. Чехова)

Syntactic transcription

(po leads-Dp 2 o 3st+5 WHICH5+2-Lp5 S 4+6-PRES4 NE6 vrem’a 6+5-Ns6 teper’ 6 govorit-INF7+3 podrobnosti-ADV3 1s1-N1 S 0+1 dolzhen1+8-SH1-sm1 by-PAST0-sm1 postupil-INF8+1_v1+9 lakejsko-Np9 k 1+10 odkaz10-Dsm10 peterburgsko10-Dsm10 chinvnik10-Ds10 po10-11 famili11-Ds11 Orlov10-Ds10 S 0+1 by1+2-PAST0-sm1 32-Dsm2 okolo1 tricat’1-G1 p’at’1-G1 let1-Gp1 i0 S 0+3 zvav3+2-PAST0-p3 32- Asm2 Georgij2-Ivanyč2-Iz2 .

k 1+2 ET2-Dsm2 Orlov2-Ds2 S 0+1 postupil-PAST0-sm1 1s1-N1 radi1+3 32-Gsm2 ot3-Gs3 izvestn3-Gsm3 gosudarstven3-Gs3 čelovek3-Gs3 WHICH5+3-Asm5 S 4+1 sčita1j5+5-PAST4-sm1 1s1-N1 serzhin5-Is5 vrag5-Is5 SE1-POSS6+1-Gs6 del6-Gs6 1s1-N1 S 0+1 rassčet’vaj1+2-PAST0-sm1 k2 živ1- GER1 u1+3 syn3-Gs3 po1+4 razgovor4-Dp4 WHICH6+4-Ap6 S 5+1 uslyša1+6-PRES5-1s1 i1 po1+7 bumag3-Dp7 l7 zapisk7-Dp7 K 9+7-SUCH5-Ap9 S 8+1 bud1-1s1 naxodi1+9-INF8+1 na9-10 stol10-Ls10 1s1-N1 v1+11 podrobnost’11-Ap11 S 2+1 izuči1+12-PRES2-1s1 plan12-Ap12 l12 namerenj12-Ap12 ot3-Gs13 .

obyknovenn1-ADV1 čas2-Gp2 v1+2 odinnadcat’2-A2 utr3-Gs3 v1+4 1s5-POSS4+5-Lsf4 lakejsko4-Ls4 S 0+1 treščat’1-PAST0-sm1 električesk1-Nsm1 zvonk1-Ns1 davaj1+6+5-GER1 1s5-Ds5 znaj5+7-INF6+5 k7 S 7+8 prosnu8-PAST7-
sm8-SE8 barin8-Ns8 . K3+1-TIME3 ls3-N3 s3+4 vyčisti4-PPT4-Isn4 platj4-Is4 i4 sapog4-Ip4 S2+3 prixodi-PAST2-sm3 v3+5 spal’n’5-As5 Georgij1-N1 Ivanyčič-N1 i1-1 TIME1 aj1-1 PAPT1-Nsm1 a1-1 skor1-1 COMP1-ADV1 utomi1-1+1-PPT1-Nsm1 snj-Is7 i0 S0+1 gl’ađe1-PAST0-sm1 v1+8 odn8-Asf8 točk8-As8 NE1 vykazij1+9-GER1 po1+10 povodi10-DS10 SE1-P0S11+1-Gsn11 probuždenij11-Gs11 NE9-K9-SUCH9-Gsn9 u dovol’stvij9-Gsn9 . 1s1+N1 S0+1 pomaga1+2+3-PAST0-sm1 32-Dsm2 odevaj2-INF3+2SE2 a0 32-Nsm2 NE2-oxotn2-ADV2 S0+2 podčin’aj2+1-PAST0-sm2-SE2 1s1-D1 molča2-GER2 i2 NE2 zamečaj2+4-GER2 1s1-POS54+1-Gsn9 prisutstvij4-Gs4 . potom1 s1+2 mokr2-Isf2 ot2+3 umyvanj-Gs3 golov2-Is2 i1 paxnu1-APT1+4-Nsm1 svež4-Ip4 dux4-Ip4 31-Nsm1 S0+1 id1+5-PAST0-sm1 v1+6 stolov6-As6 pji1+3-INF5+1 kofe9-2. 31-Nsm1 S0+1 side1-PAST0-sm1 za1+2 stol2-Is2 S0+1 pji1+3-PAST0-sm1 kofe9-A3 i0 S0+1 perelystij1+4-PAST0-sm1 gazet4-Ap4 a0 1s5-N5 is gorničn8-NSn5 Pol’5-N5 počet’l’n5-ADV5 S0+5 stojå3-PAST0-pol u5+6 dver’-Gsn6 i0 S0+5 smotres5-PAST0-pol na5+1 31-Asm1 . dv1+N1 vzrsl1-Gpi čelovek1-Gs1 S0+1 dolžn1+2-SH1-p1 by1-PAST0-p1 S0+3 sam3-Isn3 serjozn3-Isn3 vnimanij3-Is3 smotre1+4-INF2+2 K4-SUCH4 tretj-Nsn1 S0+1 pji1+5-PRES4-3S1 kofe9-A5 i4 S4+1 gryz1+6-PRES4-3S1 suxarak6-Ap6 . ET1-Nsn1 po1+2 vs’2-DSf2 verojatnost’2-DS2 S0+1-PRES0 smešn1-SH1-sm1 i1 dik1-SH1-sm1 no0 1s3-N3 S0+3 NE3 vide3+4-PAST0-sm3 dl’a3+3 SE3-Gs3 NE4-K4-Gsn4 unizitel’n4-Gsn4 v4+5 T3-Lsn5 Ks S5+6 prixodi6-PAST5-sm6-SE6 stojå3-INF6+3 okolo3+7 dver’-Gs7 xot’a3+6 S8+3 by3-PAST8-sm3 T3-SUCH3-Ism3 žé3 dvor’ánin3-Is3 i3 obrazovaje3-PPT3-Ism3 čelovek3-Is3 K9+3-SUCH0 sam0-Nsm9 Orlov9-N9 .

\[ u1+2 \text{I} 2-G2 T1-TIME1 S0+1 načinaj1-PAST0-sf1-SE1 čaxotk1-Ns1 a0 S0+3 s3+1 31-Isf1 ječoš0 IND3-K3+Nn3 požalj1+2-PRES0-Is1 pod2+3 vlijanij3-Is3 li2 bolezn’4-Gs4 ili2 načinaj5-PAST5-APT5-Gsf5-SE5 perenem5-Gs5 mirovozrenij6-Gs6 WHICH8+5-Gs8 Is1-N1 T1-TIME1 S7+1 NE1 zamečaj1+8-PAST7-sm1 1s1-Is1 iz9+10 dn’-10-Gs10 v9+11 dn’-11-As11 S2+9 ovladevaj9+1-PAST0-sf0 strast9-Nsf9 razdraž9-APT9-Nsf9 žažd9-Nsf9 obyknoven112-Gsf12 obyvatel’sk12-Gsf12 živn12-Gsf12 . 1s2-D2 S0+1 xote1+2+3-PAST0-sm1-SE1 duševn3-Gsm3 pokoj3-Gs3 zodorovj3-Gs3 xoroš3-Gsm3 vozdu3-Gs3 sytost3-Gs3 . Is1-N1 S0+1 stanov1+1-PAST0-sm1-SE1 mečtate1’-Is1 i0 K1-SUCH1 mečtate1’-Ns1 S0+1 NE1 zna1+2-PAST0-sm1 K3-Nn3 sobstvenn3-ADV3 Is1-D1 S2+3-PRES2 nužn3+1-SH3-sm3 .\]

References


2006  *Semiotaxis*. Amsterdam: UP.

Joos, Martin

1957  *Readings in linguistics I*. Chicago: UP.

Kortlandt, Frederik


