Rule-Based Dependency Parser Refined by Empirical and Corpus Statistics

Igor Boguslavsky
Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow / Universidad Politécnica de Madrid
igor.m.boguslavsky@gmail.com

Leonid Iomdin
Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow
iomdin@iitp.ru

Victor Sizov
Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow
sizov@iitp.ru

Leonid Tsinman
Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow
cinman@iitp.ru

Vadim Petrochenkov
Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow
vadim.petrochenkov@gmail.com

Abstract

The paper presents a large-coverage rule-based dependency parser for Russian, ETAP-3, and results of its evaluation according to several criteria.

The parser takes a morphological structure of a sentence processed as input and builds a dependency tree for this sentence using a set of syntactic rules. Each rule establishes one labeled and directed link between two words of a sentence that form a specific syntactic construction. The parser makes use of about 65 different syntactic links. The rules are applied by an algorithm that at first builds all possible hypothetical links and then uses a variety of filters to delete excessive links so that the remaining ones form a dependency tree. Several types of data collected either empirically or from a syntactically tagged corpus of Russian, SynTagRus, are used at this filtering stage to refine the parser performance.

The parser utilizes a highly structured 120,000-strong Russian dictionary, whose entries contain detailed descriptions of syntactic, semantic and other properties of words. A notable proportion of the links in the output trees are non-projective.

An important feature of the parser is its ability to produce multiple parses for the same sentence. In a special mode of operation, the parser may be instructed to produce more parsing outputs in addition to the first one. This can be done automatically or interactively.

In the evaluation, SynTagRus is viewed as a gold standard. Evaluation results show the figures of 0.900 for unlabelled attachment score, 0.860 for labeled attachment score, and 0.492 for unlabeled structure correctness.

1 Introductory Remarks

The syntactic parser, developed by a research team of the Institute for Information Transmission Problems in Moscow for a multipurpose linguistic processor, ETAP-3 (see e.g. Apresjan *et al.* 2003) is in many respects based on the general linguistic framework of the Meaning ⇔ Text theory, proposed by Igor Mel'čuk (e.g. Mel'čuk 1974) – especially the syntactic component of this theory. The parser is fully operational for two languages: English and Russian.

In this paper, the Russian option of the parser will be considered. Within the ETAP-3 linguistic processor, it is used in a number of applications, including Russian-to-English machine translation and the tagger for the syntactic annotation of a Russian text corpus SynTagRus.

2 Morphological Analyzer

During text analysis, the parser proper operates after the **morphological analyzer** has processed the text sentence by sentence and produced a **morphological structure** (MorphS) for each sentence. MorphS is the ordered sequence of all words of a sentence, each one represented by a lemma name, a POS attribute and a set of morphological features. The morphological analyzer works, essentially, with individual words, with a relatively few number of cases where a collocation (like *vse ravno* 'all the same') or a compound preposition (like *so storony* 'on the part of') are viewed as indivisible words. If a word form is lexically and/or morphologically ambiguous, it appears in the MorphS as a set of objects, somewhat loosely called **homonyms**, each consisting again of a lemma name, a POS attribute and a set of morphological features.

To give an example, the sentence

(1) *Inostrannye rabočie často ploxo znajut russkij jazyk* (lit. *foreign workers often badly know Russian language*) 'Foreign workers often have a poor knowledge of Russian'

will yield the following MorphS:

1.1 INOSTRANNYJ	A,NOM,PL
1.2 INOSTRANNYJ	A,ACC,INANIM,PL
2.1 RABOČIJ1	A,NOM,PL
2.2 RABOČIJ1	A,ACC,INANIM,PL
2.3 RABOČIJ2	N,NOM,PL,MASC,ANIM
3.1 ČASTYJ	A,SG,SHORT,NEUT
3.2 ČASTO	ADV
4.1 PLOXOJ	A,SG,SHORT,NEUT
4.2 PLOXO	ADV
5.1 ZNAT'	V,NONPAST,NONPERF,PL,3P
6.1 RUSSKIJ1	A,NOM,SG,MASC
6.2 RUSSKIJ1	A,ACC,INANIM,SG,MASC
6.3 RUSSKIJ2	N,NOM,SG,MASC,INANIM
7.1 JAZYK1	N,NOM,SG,MASC,INANIM
7.2 JAZYK1	N,ACC,SG,MASC,INANIM
7.3 JAZYK2	N,NOM,SG,MASC,INANIM
7.4 JAZYK2	N,ACC,SG,MASC,INANIM
7.5 JAZYK3	N,NOM,SG,MASC,ANIM

Here, A, ADV, N, and V denote, respectively, the adjective, adverb, noun and verb; NOM and ACC stand for the nominative and the accusative cases; SG and PL mark the singular and plural numbers. MASC and NEUT denote the masculine and the neutral gender. SHORT represents the short form of the adjective. ANIM and INANIM represent the animateness/inanimateness of adjectives and nouns. NONPAST, NONPERF and 3P show the present tense, the imperfective aspect and the third person of the verb.

As it happens, all words of (1) except word 5 (the verb 'know') are ambiguous. In particular, word 6 is lexically ambiguous between adjective 'Russian' and noun 'the Russian', both varying in case marking; words 3 and 4 may both be interpreted as adverbs ('often', 'badly') or adjectives ('frequent', 'bad'), whilst word 7 has three lexical readings corresponding to 'language', 'tongue', and 'prisoner', of which the former two, being inanimate, have the same forms for the nominative and the accusative case.

Accordingly, (1) consisting of 7 words has a MorphS that has as many as 18 homonyms.

The morphological analyzer is based on a comprehensive morphological dictionary of Russian that counts about 130,000 entries (over 4 million word forms).

ETAP-3 parser does not have a separate POS tagger; however, there is a small post-morphological module that partially resolves lexical and morphological ambiguity taking account of near linear context. In the case of sentence (1), this module will only delete 2 homonyms and reduce the strength of one more. On average, the module purges about 20% of homonyms.

3 The Parser

3.1 Parser Essentials

The syntactic analyzer takes a MorphS of a sentence processed as input and builds a dependency tree for this sentence using a set of syntactic rules, or **syntagms**. Each syntagm is a rule designed to establish one labeled and directed link between two words of a sentence that form a specific syntactic construction: in other words, any syntagm produces a minimal subtree that consists of two words and a link between them. There are 65 different syntactic links; e.g. the predicative link marks the domination by a finite verb [X] of its subject [Y], as in *John* [Y] *sees* [X]; the 1st completive link represents the relation between a predicate word as head and a word instantiating its 2nd valency as daughter, as in *sees* [X] *light* [Y] or *aware* [X] *of* [Y] (*my presence*), etc. Syntagms are used by the parsing algorithm that starts by building all possible hypothetical links and then uses a

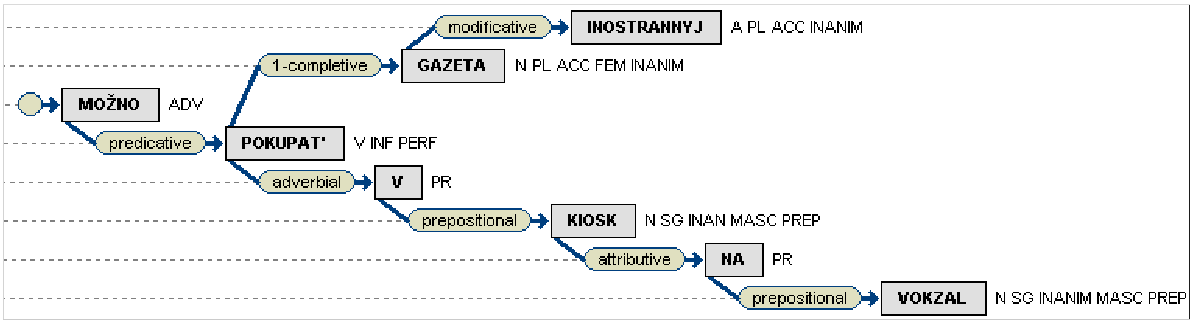

Figure 1: The dependency tree for sentence (2)

variety of filters to delete excessive links so that the remaining ones form a dependency tree.

These filters are of diverse nature and may involve data on agreement or government, repeatability/non-repeatability of specific syntactic relations (e.g. a verb may have several adverbial modifiers attached by the adverbial relation but only one subject or one direct object attached by the predicative or 1st completive relation[1]), data on link projectivity (by default, any link is projective unless a set of specific conditions are met[2]).

Fig. 1 above shows the dependency tree of the sentence

(2) *Inostrannye gazety mozhno kupit v kioske na vokzale* 'One can buy foreign newspapers at a news-stand in the railway station'.

We can see that the 1st completive link going from the verb *pokupat'* 'buy' to the noun *gazeta* 'newspaper' is non-projective as it crosses the projection of *mozhno* 'one can', which is the absolute head of the tree.

The parser makes use of a highly structured 120,000-strong Russian dictionary, whose entries contain detailed descriptions of syntactic, semantic and combinatorial properties of words.

An important feature of the parser is its ability to produce multiple parses for the same sentence. While every effort is made to ensure that the first parse obtained adequately reflects the structure of the sentence, this is not always the case. In the supervised mode of operation, the parser may be instructed to produce more parses in addition to the first one if it is unsatisfactory (the first parse may be outright wrong or, in the case of a genuinely ambiguous sentence, it may correspond to a different interpretation than that expected for the text processed) This can be done automatically or interactively, with a targeted choice of word and/or link interpretations (cf. Boguslavsky *et al.* 2005).

The parser operates in a sufficiently robust way: in the worst case, if no adequate tree can be obtained for a sentence, some of its words are linked by a soft-fail **fictitious** syntactic relation. Words that could not be found in the dictionary receive a special POS attribute **NID** (non-identified word).

Normally, each node in the resulting tree corresponds to one word of the sentence parsed. Exceptions are cases where a word is a composite not assigned a dictionary entry (such as *vos'mitomnyj* 'eight-volume), for which the parser produces two (or more) nodes in the dependency tree.

3.2 Empirical Refinement: Intersynt Duplicates of Syntagms

At the filtering stage of the analysis algorithm, two different modules can be additionally involved in order to improve the performance of the parser.

The first module (cf. Tsinman and Druzhkin 2008) is based on close empirical observation of parsed linguistic material. Basically, it implements the idea that close links between the words, especially those responsible for the core, or "skeleton", structure of the sentence, have a noticeably higher occurrence than the respective long-distance links.

In order to account for this fact, the most important syntagms (around 60 of the total number of over 200) were replicated in simpler rules that do not establish any links but increase or diminish the priority of links already established. These new rules, called Intersynt rules, work after the syntagms have

[1] In case of subject/object coordination, only one predicative or 1st completive relation is established between the predicate and the head of the coordination string (the leftmost member of this string).

[2] It turns out that even though a notable proportion of the links in dependency trees are non-projective (averagely, about 10% of processed sentences contain at least one non-projective link), the share of such links in the total amount of produced links is less than 1%.

been applied and check the bulk of the conditions specified in syntagms (disregarding some of the most subtle ones) but, unlike syntagms, operate within a very narrow space of the sentence (usually at a distance of no more than 4 words). As a result, many close links are reliably confirmed and remain in the sentence tree.

3.3 Data-Driven Statistical Refinement: Statistics of the Tagged Corpus of Texts

The data-driven statistical module (cf. Petrochenkov and Sizov 2010) collects the statistics of links from the syntactically annotated corpus of Russian texts, to be described in more detail in Section 5 below. Statistical data represent the distribution of syntactic links in the treebank that takes account of the following three factors: 1) the distance between the words, 2) the direction of the link (from left to right or vice versa), and 3) the number of the word sense of the word involved in the link (normally, lexical meanings of polysemantic words are ordered in the dictionary in such a way that the more general and more frequently used meanings have smaller numbers than the peripheral meanings). The statistical module intervenes at the moment when the parser chooses among the established competing syntactic hypotheses for an undecided syntactic daughter and prompts the algorithm to select the link occurring in the tagged corpus in similar environment with the maximum frequency.

In different modes of operation, the parser may use either of the two modules, both of them, or neither. The use of the empirical module turned out to provide a noticeable improvement to parser performance as compared to the “bare” parser.

4 The Corpus

The ETAP-3 parser is used to construct the first Russian dependency treebank, SynTagRus (Boguslavsky *et al.* 2000, 2009; Apresjan *et al.* 2006). Currently the treebank counts over 45,000 sentences (650,000 words) belonging to texts from a variety of genres (contemporary fiction, popular science, newspaper, magazine and journal articles dated between 1960 and 2011, texts of online news, etc.) and is steadily growing.

Since Russian, as other Slavic languages, has a relatively free word order, SynTagRus adopted a dependency-based annotation scheme, in some respects parallel to the Prague Dependency Treebank (Hajič *et al.*, 2001). Syntactic tagging makes use of the full list of the 65 syntactic relations active in the parser (plus one or two specially introduced relations that cannot be handled automatically). All sentences are supplied by a complete tree structure, even if the parser cannot build one. The fictitious link mentioned above is not allowed.

The corpus is built semi-automatically: first, each sentence is processed by the ETAP-3 parser, then it is manually edited by expert linguists, who correct errors made by the parser and handle cases of ambiguity that cannot be reliably resolved without extralinguistic knowledge.

During the manual stage of corpus creation, certain improvements are introduced into the dependency tree annotation that cannot be achieved automatically. In particular, hard cases of ellipsis are made explicit by introducing additional nodes into the annotation. A sentence like

(3) *Ja priexal iz Moskvy, a on iz Madrida* ‘I came from Moscow and he from Madrid’

will receive a resulting tree with another instance of the verb *priexal* ‘came’ so that the syntactic links that form the tree have a more natural look. This additional node is marked with a special phantom label.

In this study, SynTagRus is used as gold standard for parsing evaluation. As a matter of fact, the corpus has already been used for a number of linguistic research and development tasks. In particular, it was used as benchmark in regression tests designed to ensure stable performance of the ETAP-3 Russian parser in the course of its development (see e.g. Boguslavsky *et al*. 2008) and as a source for the creation, by machine learning methods, of a successful statistical parser for Russian (Nivre *et al*., 2008).

5 Evaluation Metrics

We use two types of evaluation: a general evaluation and a penalty-based one. Both will be briefly characterized below.

5.1 General evaluation

Lexico-Grammatical Score (LG)

As discussed in Section 3 above, ETAP-3 does not have a separate POS-tagging stage. Disambiguation of lexico-grammatical features is carried out in parallel with establishing dependency links. However, it is useful to evaluate the POS attribution accuracy separately. It is calculated as follows. For each identified word L its lexico-grammatical coefficient KL = n_1/n_2 is determined, where n_1 is the number of correctly identified features of L, and n_2 is the number of all its features. Lexico-grammatical score is defined as the sum of all lexico-grammatical coefficients divided by the number of words.

Word-oriented syntactic scores

- **Head Score**: proportion of words for which the head (or the absence of a head) has been assigned correctly (= Unlabelled Attachment Score; Nivre and Scholz 2004, Eisner 1996).
- **Link Score**: proportion of words for which a name of subordinating link (or the absence of the head) has been assigned correctly. This link may depart from a wrong head.
- **Head and Link Score**: proportion of words for which both the head and the label for subordinating link have been identified correctly (= Labelled Attachment Score; Lin 1998, Nivre and Scholz 2004).
- **Link Audit**: for each link type, its precision, recall and F-score are calculated.

Sentence-oriented syntactic scores

These scores can be computed either for the whole corpus, or for sentences of certain length, e.g. for sentences with less than 10 words, with 10-20 words, with 20-30 words, etc.

- **Root Score:** proportion of sentences for which the root has been identified correctly.
- **Unlabeled Structure Correctness Score**: proportion of sentences for which all the links have been identified correctly – with no regard to link labels (= Complete Rate of Yamada and Matsumoto 2003).
- **Strict Structure Correctness Score**: proportion of sentences for which all links and their labels have been identified correctly.
- **Gold Standard Achievability**: proportion of sentences for which the gold standard (GS) structure is achieved within the first N alternatives in the stack.

GS achievability is an important feature of the parser. As mentioned in 4.1, the ETAP-3 parser can produce all alternative parses compatible with the grammar. The order in which these alternatives are presented depends on the rank which the parser assigns to them. Sometimes the parser is able to obtain GS but this parse is not on the top of the stack of alternatives. It is useful to know how many GS parses the parser can produce, even if not as the first alternative. This score shows what proportion of incorrect parses is due to grammar flaws as opposed to defects that could be eliminated by means of a better ordering of alternatives. The GS achievability score provides information on the proportion of sentences which achieved GS within the first N alternatives and some other types of supplementary information.

5.2 Penalty-based evaluation

This type of evaluation is based on a detailed list of possible types of deviation of a parse (P) from the gold standard (GS). These types are as follows.

Tokenization deviations

- GS contains a phantom node which has no match in P (see Section 5 above).
- A string of characters in the sentence is differently segmented into tokens in P and GS. This happens when a multiword expression is treated as one word by the corpus annotator but not by the parser dictionary. Here two cases can be distinguished: (a) the difference is recoverable, i.e. one can automatically match nodes in P and GS, and (b) it is unrecoverable. Example of case (a): *antiterrorizm* is represented by one node in GS, but corresponds to two nodes in P (*anti* and *terrorizm*). The parser did not find *antiterrorizm* in the dictionary but decomposed it into two parts and connected them with a composite link. In this case, *antiterrorizm* in GS matches with *terrorizm* in P for further comparison. It is easy since both items have the same list of features. Example

of case (b): the annotator decides that a multiword expression should be represented in the corpus as one indivisible word, which is not present in the dictionary. In this case, the GS may contain e.g. an adverb like *po krajnej mere* ('at least') which is hardly possible to match with the sequence of preposition *po*, adjective *krajnej* and noun *mere* present in P.

Lexico-grammatical deviations between nodes in P and GS with identical tokens

- The word is not recognized in P. It is absent from the dictionary and cannot be decomposed derivationally.
- Nodes in P and GS have different parts of speech, e.g. *čto* can be a pronoun 'what' or a conjunction 'that'.
- Nodes in P and GS have different features within the same part of speech (for examples, see Section 3).
- Nodes in P and GS have different lemmas within the same part of speech, e.g. *naxodit'sja* can be interpreted either as the verb meaning 'be located (somewhere)' or as the passive of the verb *naxodit'* 'find'.

Syntactic deviations between nodes in P and GS with identical tokens

All syntactic deviations are mutually exclusive.

- The node in P is connected to another node by a **fictitious** link.
- The node is the root in P but not in GS, or vice versa.
- The node in GS is connected by a link which is absent in the list of links supported by the parser. This may happen, since SynTagRus contains some specific constructions annotated manually.
- In GS the node is linked to node Z with relation R, and in P it is also linked to Z, but with a relation different from R.
- In GS the node is linked to node Z with relation R, and in P it is also linked with R, but to a node different from Z.
- In GS the node is linked to node Z with relation R, and in P it is neither linked to node Z, nor with relation R.

Each deviation type is assigned a penalty. Accordingly, we can calculate penalties of nodes, parses of sentences and parses of corpora. Two types of evaluation can be used.

Non-normalized evaluation is very simple and convenient for comparing results obtained on the same corpus at different times. It consists in summing up all penalties assigned in parsing the corpus. **Normalized evaluation** permits to compare the results obtained on different corpora. It is calculated as follows. For each node, its penalties are summed up and divided by the maximum penalty a node can get. For a sentence, node evaluations are summed up and divided by the number of nodes composing the sentence. For a corpus, sentence evaluations are summed up and divided by a number of sentences in the corpus.

Besides generating the general penalty for a node, sentence or corpus, one can identify a number of specific errors, which helps parser developers to assess the processing accuracy for certain syntactic phenomena. Among them, failures can be detected in:

- finding actants of finite verbs, non-finite verbs, nouns, adjectives and adverbs,
- finding the subject of a verbless sentence,
- finding non-actant dependents of verbs,
- establishing various types of auxiliary links,
- identifying coordination chains.

The syntactic model underlying the parser includes several weakly contrasting dependency types, e.g. different types of attributes and modifiers. One could think of merging them into one hyper-dependency type so as to increase the accuracy of the model. The evaluation software provides a convenient tool to assess the effect of such a merge without the need to previously introduce complex changes to the rules. Specifically, the program can be instructed to disregard certain types of syntactic deviations. For example, one can evaluate the parse of the corpus under the condition that relations R1 and R2 are identical.

6 ETAP-3 Parser Evaluation

Below, some general evaluation data obtained on a fragment of the SynTagRus corpus are presented. This fragment is selected relatively randomly: it represents complete data introduced in the corpus in 2007. The fragment contains 66401 words in 4676 sentences. We will give the results of two types of evaluation: **strict evaluation,** which involves straightforward calculation of the parameters listed in Section 6, and **relaxed evaluation**, which ignores certain deviations between the gold standard and the evaluated performance of the parser.

The evaluation is largely based on the data from the same syntactically tagged Russian corpus which is used for parser refinement. Methodologically, this is in our opinion quite acceptable, since the version that we evaluated does not include any machine learning.

6.1 Strict evaluation

As was mentioned in section 4, the parser has several modes of operation, including (a) the default mode using the empirical module (EM), (b) the mode that incorporates a data-driven statistical component (DD mode) and (c) the mode that uses neither of the two. The DD component was trained on a different fragment of SynTagRus than that used for evaluation: it includes the data introduced in 2009 (7379 sentences with 103694 words).

Our experiments show that mode (a) where the EM component is used yields the best quality: we will treat it as the default mode. The DD mode yields a slightly worse parsing quality but operates substantially faster. The results obtained in the default mode are given below.

Lexico-Gram. Score	0.977
Head Score (UAS)	0.900
Link Score	0.887
Head&Link Score (LAS)	0.860
Unlabeled Struct. Correctness	0.492
Strict Struct. Correctness	0.352
GS Achievability (stack of 5)	0.511

Table 1. Strict evaluation for the default mode

The Gold Standard Achievability within the first 5 trees in the stack reaches 0.512. This figure is worth comparing with the Strict Structure Correctness score. While the first tree in the stack coincides with the Gold Standard in 35.2% of cases, the Gold Standard tree is found among the first 5 trees in the stack in 51.1% of cases.

The table below presents the link audit calculated on the first alternative basis.

LINK NAME	RECALL	PREC.	F-SC.
1- complement	0.895	0.900	0.897
2- complement	0.815	0.747	0.780
3- complement	0.738	0.629	0.679
4- complement	0.667	0.267	0.380
Appositive	0.855	0.820	0.838
Attributive	0.713	0.631	0.670
Parenthetical	0.832	0.903	0.866
Durative	0.638	0.620	0.673
Infinitive-conjunctive	0.955	0.984	0.969
Quasiagentive	0.927	0.877	0.901
Quantitative	0.938	0.956	0.947
Nonactant-completive	0.741	0.642	0.688
Circumstantial	0.732	0.881	0.800
Restrictive	0.936	0.872	0.903
Modificative	0.966	0.984	0.975
Passive-analytical	0.986	0.973	0.979
Subordinative-conjunctive	0.863	0.867	0.865
Predicative	0.906	0.941	0.923
Prepositional	0.985	0.990	0.988
Copulative	0.858	0.895	0.876
Proleptic	0.475	0.848	0.609
Explicative	0.744	0.668	0.703
Relative	0.830	0.899	0.863
Sentential-coordinative	0.724	0.601	0.657
Coordinative-conjunctive	0.877	0.909	0.893
Coordinative	0.864	0.875	0.869
Comparative-conjunctive	0.809	0.793	0.800
Comparative	0.859	0.751	0.801
Expletive	0.841	0.860	0.850
Elective	0.868	0.951	0.908

Table 4: Link audit

6.2 Relaxed evaluation

In this type of evaluation, the comparison criteria for the parser and the gold standard were weakened as follows.

1) Certain poorly distinguishable syntactic links that the annotators failed to treat in a consistent way throughout the corpus were considered as one link. This was e.g. the case with three types of links that represent different kinds of apposition (the appositive, the nominative-appositive, and the numerative-appositive links: prototypical examples are, respectively, Russian equivalents of phrases like *President Medvedev*, *Novel 'Gone with the wind'* and *Group Three*. Other link clusters included (a) the parenthetical and the restrictive relation for cases like *In particular, they refused to obey* vs. *They refused to obey, in particular John*, and (b) agentive and 2nd

completive relations for cases such as *On byl ubit otravlennoj streloj* 'He was killed by/with a poisoned arrow': in one interpretation, the arrow is the agent whilst in the other it is the tool.

2) Certain differences between the parses were ignored if the correct choice required deep semantic knowledge. Primarily, this was the case with different PP attachment in sentences like *He saw a girl with a telescope.*

The following data are the results of relaxed evaluation for the default parsing mode.

	Relaxed eval. default	Wrt Strict Eval.
Lexico-Gram. Score	0.978	+0.001
Head Score (UAS)	0.918	+0.018
Link Score	0.904	+0.017
Head&Link Score (LAS)	0.885	+0.025
Unlabeled Str. Correct.	0.582	+0.090
Strict Struct. Correctness	0.439	+0.087
GS Achievability (stack of 5)	0.560	+0.049

Table 5: Relaxed evaluation for the default mode

The most notable distinction from the strict evaluation is the increase of the Head & Link Score by 2.5%, as well as the increase of the Unlabeled structure and the Strict Structure Correctness by 9.0% and 8.7%, respectively. GS Achievability also grew by 4.9%.

6.3 Comparison with related work

To the best of our knowledge, there are no data on Russian parsers with which we could compare our results. The only exception is the data-driven MaltParser by J. Nivre trained on the SynTagRus corpus (Nivre *et al.* 2008). That is, both parsers strive to come to exactly the same structures, which provides favorable conditions for comparison. However, direct collation of the ETAP-3 parser performance with the results obtained by J. Nivre would hardly be correct, since the input of these parsers is significantly different. The ETAP-3 parser processes the raw, unprepared text. The MaltParser begins with the POS-tagger output. Since no such tagger for Russian was available for the experiments, the input was taken directly from the GS. This means that all tokenization and lexico-grammatical deviations between the sentence and GS (cf. 6.2 above) have been rid of in advance. It is difficult to accurately assess the impact of these deviations on the ETAP-3 performance. This being said, one can compare two scores available for both parsers. They are largely similar: Head Score – 0,900 (ETAP-3, strict evaluation mode) vs. 0.891 (MaltParser), Head and Link Score – 0.860 (ETAP-3, strict evaluation mode) vs. 0.823 (MaltParser).

As for the related work on dependency parsers for other languages, we can compare the Unlabeled Structure Correctness of ETAP with the English data in Collins 1997, Charniak 2000, Yamata and Matsumoto 2003 and in Nivre and Scholz 2004:

Charniak	0.452
Collins	0.433
Yamada & Matsumoto	0.384
Nivre & Scholz	0.304
ETAP-3 (strict evaluation)	0.492

Table 6: Unlabeled Structure Correctness Score

Additionally, our Head & Link score (both for strict and relaxed evaluation) proves visibly higher than the average figure for this parameter (0.8253) given for several languages in Nivre and McDonald 2008.

7 Error Analysis

As seen from Table 4, of 30 dependency relations represented in the corpus, there are 8 whose F-score exceeds 0.9, and 8 stay below 0.7. We will illustrate both groups of relations with short examples (the head of the construction will be denoted in the gloss as X and the subordinate as Y).

High accuracy relations: infinitival-conjunctive (*čtoby vstretit'* 'in-order-to [X] meet [Y]'), restrictive (*ne byl* 'was not', lit. 'not [Y] was [X]'), quantitative (*pjat' dnej* 'five [Y=Nom] days [X=Gen]'), modificative (*tri opytnyx rabotnika* 'three experienced [Y=Pl] workers [X=Sg]'), passive-analytical (*byl isključen* 'was [X] expelled [Y]'), predicative (*solnce svetit* 'the sun [X] shines [Y]'), prepositional (*v dlinnom spiske* 'in [X] the long list [Y]'), elective (*samaja interesnaja iz knig* 'the most interesting [X] of [Y] the books').

Low-accuracy relations: 3-rd completive (*oprobovat' preparat na myshax* 'to test [X] the medication on [Y] mice'); 4-th completive (*arendovat' na tri goda* 'rent [X] for [Y] three years', *perevozit' počtu samoletom* 'to transport [X] mail by [Y] airplane'), attributive (*dom za uglom* 'a house [X] round [Y] the corner'), durative (*on spit po pjat' chasov v sutki* 'he sleeps [X] five hours [Y] a day'), nonactant-completive (*prishel ko mne v kabinet* lit. 'came [X] to [Y] me into my study' proleptic (*somnenija, oni dolžny byt*' 'doubts [X], they [Y] should exist'), explicative (*my kupili vse – xleb, syr, moloko* 'we bought everything [X] – bread [Y], cheese, milk'; sentential-coordinative (*Oni ne pridut, i my ostanemsja odni* 'they will [X] not come, and [Y] we will be alone').

A detailed error analysis cannot be done within a short paper. By way of example, we will only comment on the attributive link which connects a noun with its non-argument modifier if they do not agree in case, number and gender. This is a notoriously difficult link to establish, due to the absence of formal features and the abundance of possible heads. Most of the situations in which an attributive link is established erroneously are the following:

(a) it is established instead of a circumstantial link leading from a verb,

(b) it is established instead of an attributive link leading from a more distant noun,

(c) it is established instead of an appositive link, if the subordinate node is a non-identified (NID) proper noun absent in the dictionary.

The latter case deserves a special comment. Existence of NIDs significantly decreases the recall of the attributive and the precision of appositive links. This may be improved by including a Named Entity Recognizer at the preprocessing stage. Another direction of improvement is connected with augmenting the performance of the guessing rules which should identify the morphological form of the word even if it is absent from the dictionary.

One more notable source of parser failure is inconsistent dictionary coverage. In many cases, ignorance is better than half-truth: it is better to leave a whole family of lexical units outside the dictionary than to introduce it fragmentarily. Consider a typical situation where a Russian name of a town, like *Krasnojarsk*, is present in the dictionary but the corresponding adjective, *krasnojarskij*, is not. Due to a specific intersection of paradigms of such words (they have coinciding word forms in structurally different cases: the instrumental case of the noun coincides with the locative case of the adjective), sentences like

(3) *On rabotaet na krasnojarskom zavode* 'He works at a Krasnoyarsk plant'

will not be parsed sensibly because the adjective will be treated as a stray noun in the instrumental case. It would be counterintuitive to instruct the parser to treat a word form found in the dictionary on a par with a non-identified word. Accordingly, the parse would be more acceptable if the whole family of words remained unlisted: in this case, there will be a local parsing mistake, whereas in the opposite case the parser will simply play havoc with the structure.

8 Conclusion

We have presented ETAP-3 parser, a rule-based system for dependency parsing which makes part of a multifunctional linguistic processor. It was developed for Russian and English, but evaluated only for Russian by means of a SynTagRus dependency treebank. The characteristic feature of the parser is a fine-grained dependency type set which includes 65 types. Some of them are rather rare: in the fragment of the treebank used for evaluation, only 30 types are represented. We use various types of metrics, some of them better suited for intrinsic evaluation (penalty-based), while others (general) are convenient for comparison with other systems.

The main directions of future research are: 1) improvement of rules for low score dependency types, 2) development of rules for treating ellipsis, 3) upgrading the algorithm for producing alternative parses, 4) experiments on developing a hybrid rule-based/data-driven parser.

9 Acknowledgements

This study has been partially funded by the Russian Foundation for Basic Research (grants No. 10-07-90001-Bel_a and 11-06-00405-a, and a 2011 Grant on Corpus Linguistics from the Russian Academy of Sciences. The authors express their appreciation to the Foundation and the Academy.

References

Jurij Apresian, Igor Boguslavsky, Leonid Iomdin, Alexander Lazursky, Vladimir Sannikov, Victor Sizov, Leonid Tsinman. 2003. ETAP-3 Linguistic Processor: a Full-Fledged NLP

Implementation of the MTT. // MTT 2003, First International Conference on Meaning – Text Theory. Paris, École Normale Supérieure, Paris, pp. 279-288.

Juri Apresjan, Igor Boguslavsky, Boris Iomdin, Leonid Iomdin, Andrei Sannikov, Victor Sizov. (2006). A Syntactically and Semantically Tagged Corpus of Russian: State of the Art and Prospects.// Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006). Genoa, p. 1378-1381.

Igor Boguslavsky, Svetlana Grigorieva, Nikolai Grigoriev, Leonid Kreidlin, Nadezhda Frid. 2000. Dependency Treebank for Russian: Concept, Tools, Types of Information // Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000), p. 987-991)

Igor M. Boguslavsky, Leonid L. Iomdin et al. (2005). Interactive Resolution of Intrinsic and Translational Ambiguity in a Machine Translation System. // CICLing 2005. Lecture notes in computer science. A.Gelbukh (ed.), Springer-Verlag Berlin Heidelberg 2005, pp. 383 – 394.

Igor M. Boguslavsky, Leonid L. Iomdin, Svetlana P. Timoshenko, Tatyana I. Frolova (2009). Development of the Russian Tagged Corpus with Lexical and Functional Annotation // Metalanguage and Encoding Scheme Design for Digital Lexicography. MONDILEX Third Open Workshop. Proceedings. Bratislava, Slovakia. April 15-16, 2009. Bratislava, 2009. P. 83-90. ISBN 978-80-7399-745-8.

Charniak, E. 2000. A maximum-entropy-inspired parser. In Proceedings of NAACL.

Collins, M. 1997. Three generative, lexicalized models for statistical parsing. In Proceedings of ACL, pp. 16-23, Madrid.

Eisner, J.M. 1996. Three new probabilistic models for dependency parsing: An exploration. In Proceedings of COLING, Copenhagen, Danemark.

Hajič et al., 2001: Hajič, J., Vidova Hladka, B., Panevová, J. E. Hajičová, P. Sgall, and P. Pajas. 2001. Prague Dependency Treebank 1.0. LDC, 2001T10.

Lin, D. 1998. A dependency-based method for evaluating broad coverage parsers. Natural Language Engineering 4, 97-114.

Mel'čuk I.A. 1974. Opyt teorii lingvisticheskix modelej klassa "Smysl - Tekst". [The theory of linguistic models of the Meaning – Text Type]. Moscow, Nauka. (In Russian).

Nivre, J. and Scholz, M. 2004. Deterministic dependency parsing of English text. // Proceedings of the 20th International Conference on Computational Linguistics (COLING), pp. 64-70.

Nivre, J. and McDonald, R. 2008. Integrating Graph-Based and Transition-Based Dependency Parsers. In Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-08: HLT), pp. 950-958.

Petrochenkov V.V. and Sizov V.G. Ispol'zovanie statističeskoj informacii o konkurirujuščix sintaksičeskix svjazjax v sintaksičeskom analizatore ETAP-3 dlja polučenija naibolee verojatnoj sintaksičeskoj struktury frazy // Informacionnye texnologii i sistemy (ITiS'10). Sbornik trudov 33 Konferencii molodyx učenyx i specialistov IPPI RAN. Gelendžhik, 18-26 sentjabrja 2010 g. Moskva: IPPI, 2010. P. 299-305.

Tsinman, L. and Druzhkin K. (2008). Sintaksičeskij analizator lingvističeskogo processora ETAP-3: eksperimenty po ranžirovaniju sintaksičeskix gipotez // Kompjuternaja lingvistika i intellektal'nye tekhnologii (Dialog 2008). Trudy meždunarodnoj konferencii. Bekasovo, 4-8 ijunja 2008 g. Moskva, RGGU, 2008. Vyp. 7(14). P. 147-153. ISBN 978-5-7281-1022-4.

Yamada H., Matsumotu Y. 2003. Statistical dependency analysis with support vector machines. // Proceedings of the 8th International Workshop on Parsing Technologies (IWPT), pp.195-206, Nancy, France.