

## AGREEMENT IN CZECH AS MANIFESTED IN THE CZECH NATIONAL CORPUS<sup>1</sup>

### 1. Introduction

Agreement is a prominent surface syntax phenomenon and its full exploitation can substantially improve automatic analysis of Czech (especially part-of-speech and morphological disambiguation, partial and full-fledged parsing) and its synthesis. In Czech with a free word order, agreement plays a major role in specifying sentence structure. Thus, this crucial syntactic and morphological phenomenon should be paid due attention.

Agreement in Czech is a very complex phenomenon with many ramifications and special cases. This article can present only a survey of major kinds of agreement in contemporary Czech as identified in the texts collected in three synchronic corpora of Czech: SYN2000<sup>2</sup> (100 mil. running words), SYN2005<sup>3</sup> (100 mil. running words), SYN2006PUB<sup>4</sup> (300 mil. running words).

Agreement is a surface syntax phenomenon consisting in that a word form *A* and a word form *Y* in a sentence have the same value of the following semes and/or their combinations:

- person (*1st, 2nd, 3rd*) – expressed by verbs and personal pronouns in Czech;

---

<sup>1</sup> The paper was supported by the grant No. MSM0021620823.

<sup>2</sup> *Czech National Corpus*. SYN2000. Prague, 2000. Available on-line from <http://ucnk.ff.cuni.cz>

<sup>3</sup> *Czech National Corpus*. SYN2005. Prague, 2005. Available on-line from <http://ucnk.ff.cuni.cz>

<sup>4</sup> *Czech National Corpus*. SYN2006PUB. Prague, 2006. Available on-line from <http://ucnk.ff.cuni.cz>

- number (*singular, plural, rests of dual*) – expressed by nominal elements and verbs;
- gender (*masculine animate, masculine inanimate, feminine, neuter*) – expressed by nominal elements, past participles, passive participles and transgressives
- case (*nominative, genitive, dative, accusative, vocative, locative, instrumental*) – expressed by nominal elements and passive participles.

Moreover, one of the following conditions holds:

**A.** Word forms *X* and *Y* form a syntagm:

**A1.** agreement of a finite verbal predicate *X* with a corresponding subject *Y* in *number, gender* and *person*;

**A2.** agreement of a finite verbal predicate *X* with a subject complement or object complement *Y* in *number, gender* and *case*;

**A3.** agreement of a syntactically adjectival attribute *X* with the governing noun *Y* in *number, gender* and *case*;

**A4.** agreement of an nominal attribute or apposition *X* with a governing noun *Y* in *case*.

**B.** Word forms *X* and *Y* are parts of a compound predicate:

**B1.** agreement of individual verbal forms in *person, number* and *gender* within an analytically formed verbal predicate

**B2.** a word form *X* is a nominal subject, a word form *Y* is a syntactically adjectival nominal predicate and word forms *X* and *Y* agree in *number, gender* and *case*;

**B3.** a word form *X* is a verbal predicate, a word form *Y* is a nominal predicate and word forms *X* and *Y* agree in *number, gender* and *person*;

**C.** Word forms  $X_1, X_2 \dots X_n$  are conjuncts, i.e. members of the same coordinated construction – they agree in *case*.

**D.** A finite form word *X* agrees with a passive participle *Y* coforming a passive infinitive in *number* and *gender*;

**E.** Word forms *X* and *Y* are coreferential:

**E1.** a word form *X* is a nominal antecedent and a word form *Y* is a relative pronoun as its coreferential consequent – *X* agrees with *Y* in *number* and *gender*;

**E2.** other cases of agreement as a means for expressing a *coreference relation*.

**F.** A word form *X* is subject of a clause, a word form *Y* is a transgressive form in a participial construction: *X* and *Y* agree in *number* and *gender*.

**G.** Agreement of finite verbal forms in coordinated different clauses in *number*, *gender* and *person*.

**H.** Special/complex cases.

In this very short article we can present only a very brief survey of the main agreement types **A** and **B1** below.

**Type A1.** Agreement of a finite verbal predicate *X* with a corresponding subject *Y* in *number*, *gender* and *person*.

**Example 1**

(1) Žádné počasí(subj,neut-sg-3rd) nezkalilo(pred,neut-sg-3rd) jeho šoféřskou vášně.

*E. lit.* No weather polluted his driver's passion.

(2) Likvidaci musím(pred,sg-1st) platit já(subj,sg-1st).

*E. lit.* The liquidation must pay I.

*E.* I must pay the liquidation.

(3) Ale lidé(subj,mascanim-pl-3rd) ho naposledy viděli(pred,mascanim-pl-3rd) v úterý večer.

*E. lit.* But the people saw him last on Tuesday evening.

Both word order variants are possible: subject preceding the corresponding predicate or vice versa.

The subject can, however, be formed by a *coordinated construction*. In this case, four important factors come into play<sup>1</sup>:

(a) coordinated nominal elements in singular require the verbal predicate to be in *plural* if individual conjuncts do not form a semantic whole. If, however, the conjuncts form one semantic whole, the predicate is in *singular*;

(b) hierarchy of person values:

1. *first* person, 2. *second* person, 3. *third* person;

(c) hierarchy of gender values:

1. *masculine animate*, 2. *masculine inanimate / feminine*, 3. *neuter*.

Specially, two neuter nouns in singular entail the predicate in plural masculine inanimate / feminine in case they do not form a semantic whole, otherwise the predicate is in neuter singular;

(d) mutual word order position of the coordinated subject and the predicate: if the predicate precedes the subject it can agree either with the first conjunct only, or with the whole coordinated subject.

### Example 2

ad (a):

(4) V tu chvíli se *otec*(conjunct<sub>1</sub>,mascanim-sg) a *syn* conjunct<sub>2</sub>,mascanim-sg) *shodli*(pred,mascanim-pl), že nemá cenu v debatě pokračovat. (The conjuncts do not form a whole.)

*E. lit.* At that moment *father* and *son* agreed that discussion is not worth continuing.

(5) Tato *lhostejnost*(conjunct<sub>1</sub>,fem-sg) a *laxnost*(conjunct<sub>2</sub>, fem-sg) *byla*(pred,fem-sg) již *zaplácena*(pred,fem-sg) milióny lidských životů. (The conjuncts form a whole.)

*E. lit.* This *indifference* and *laxity* has already been paid by millions of human lives.

---

<sup>1</sup> Šmilauer V. Novočeská skladba. Praha, 1966. P. 72.

ad (b):

(6) Tvůj *otec*(conjunct<sub>1</sub>,sg-3rd) a *já*(conjunct<sub>2</sub>, sg-1st) ho očekáváme(pred,pl-1st) vestoje.

*E. lit.* Your *father* and *I* are waiting for him standing.

(7) Od této chvíle *ty*(conjunct<sub>1</sub>,sg-2nd) a tvůj *rod*(conjunct<sub>2</sub>,sg-3rd) *nebudete*(pred,pl-2nd) hnědí jako suchá tráva.

*E. lit.* Since that moment *you* and your *family* will not be brown like dry grass.

ad (c):

(8) Ty *květiny*(conjunct<sub>1</sub>,fem-pl) a *stromy*(conjunct<sub>2</sub>,mascinan-pl) *neměly*(pred,mascinan-pl) vůbec jméno.

*E. lit.* The *flowers* and *trees* had no name.

(9) Kde *orli*(conjunct<sub>1</sub>,mascanim-pl) a *vrány*(conjunct<sub>2</sub>,fem-pl) *označovali*(pred,mascanim-pl) místo, na němž ležel oštěpem probodnutý kůň.

*E. lit.* The *eagles* and *crows* indicated a place where a horse stabbed by a spear was lying.

(10) Není náhoda, že *obžerství*(conjunct<sub>1</sub>,neut-sg) a *smilstvo*(conjunct<sub>2</sub>,neut-sg) patřily(pred,mascinan-pl) mezi sedm smrtelných hříchů. (The conjuncts do not form a whole.)

*E. lit.* It is no chance that *gluttony* and *fornication* belonged to seven deadly sins.

(11) *Dobro*(conjunct<sub>1</sub>,neut-sg) a *zlo*(conjunct<sub>2</sub>,neut-sg) *nebylo*(pred,neut-sg) absolutní, byla to měřítko Bohova štěstí. (The conjuncts form a whole.)

*E. lit.* *Good* and *evil* was not absolute, they were criteria of God's happiness.

ad (d):

(12) V pondělí 5. září večer *byly*(pred, mascinan-pl) *kufřík*(conjunct<sub>1</sub>,mascinan-sg) a *balíček*(conjunct<sub>2</sub>, mascinan-sg) *převezeny*(pred, mascinan-pl) z banky do laboratoře MI5.

*E. lit.* On Monday 5th September evening *were* the *suitcase* and the *parcel transported* from the bank to MI5 laboratory.

(13) A tím *začala*(pred, fem-sg) *turistika*(conjunct<sub>1</sub>, fem-sg) a *přítliv*(conjunct<sub>2</sub>, masculin-sg) lidí. (The predicate verb agrees with the first conjunct.)

*E. lit.* And by that *began tourism* and *influx* of people.

**Type A2.** Agreement of a finite verbal predicate *X* with a subject or object complement *Y* in *number, gender* and *case*.

**The subject complement** is formed by a long syntactic adjective or by a short adjective or a passive participle and it is in the *nominative case*.

### Example 3

(14) *Ona*(subj, fem-sg-nom) *chodila*(pred, fem-sg) celý život *bosa/bosá*(compl, fem-sg-nom).

*E. lit.* *She was walking* the whole life *barefooted*.

(15) *Rádi*(compl, masculin-pl-nom) se *oblékali*(pred, masculin-pl) do šatů a bot z jelení kůže.

*E. lit.* They *were fond* of *wearing* the clothes and shoes from the deer skin.

**The object complement** is also formed by a long syntactic adjective or by a short adjective or a passive participle and it is in the *accusative case*:

### Example 4

(16) Kdo nebude mít *zaplacenou*(compl, fem-sg-acc) *pokutu*(obj, fem-sg-acc), tomu podnik jízdenku neprodá.

*E. lit.* Whoever will not have *paid* the *fine*, the company will not sell him the ticket.

(17) Roger nepovažoval tuto *poznámku*(obj, fem-sg-acc) za *hodnu/hodnou*(compl, fem-sg-acc) odpovědi.

*E. lit.* Roger did not consider this *remark* as *worth* replying.

Rarely, the object complement can be in an indirect case (genitive, dative, locative, instrumental) but in this case it can be formed by a long syntactic adjective only.

**Type A3.** Agreement of a syntactically adjectival attribute *X* with the governing noun *Y* in *number*, *gender* and *case*.

The syntagm (*X Y*) of this kind is the far most frequent in Czech as the corpora of Czech confirm: almost 8 % pairs in them are formed by this kind of syntagm. Both word order positions are possible:

A syntactically adjectival attribute *X* precedes its governing noun *Y*:

**Example 5**

(16) *Senzační*(attr,neut-sg-acc) *odhalení*(noun,neut-sg-acc) přinesl v těchto dnech italský tisk.

*E. lit.* A *sensational revelation* brought in these days Italian press.

*E.* A sensational revelation was published in Italian press these days.

A syntactically adjectival attribute *X* follows its governing noun *Y*:

**Example 6**

(17) *Vilka* stojí na klidném místě, ve *čtvrti*(noun,fem-sg-loc) *zaklíněné*(attr, fem-sg-loc) mezi kopce.

*E. lit.* The small villa stands on a quiet place, in a *quarter wedged* in between the hills.

**Type A4.** Agreement of an nominal attribute or apposition *X* with a governing noun *Y* in case.

**Type B1.** Elements of an analytically formed verbal predicate agree in *person*, *number* and *gender*. In Czech there are the following tenses, moods and voices:

- tenses: *preterite*, *pluperfect*, *present*, *future*;
- moods: *indicative*, *imperative*, *conditional*, *infinitive*, *transgressive*;

- voices: *active, passive*.

Some of the tense-mood-voice combinations are formed analytically (number of verbal forms  $\geq 1$ ), the other being formed synthetically. The analytically formed meanings are as follows:

**B1.1.** indicative preterite active of the 1st and 2nd person singular and plural – agreement in *number*;

**B1.2.** conditional active (all combinations of person, number, gender, voice) – agreement in *number*;

**B1.3.** indicative pluperfect active (all combinations of person, number, gender, voice) – almost missing in contemporary Czech;

**B1.4.** indicative periphrastic future active (restricted to imperfective verbs) by means of the forms of the auxiliary *být* plus active infinitive – no agreement;

**B1.5.** indicative periphrastic passive (all combinations of mood, tense, person, number, gender) – agreement in *number* and *gender*.

### Conclusion

A thorough description of all kinds of agreement in Czech corpora with respect to word order and other syntagmatic aspects is inevitable for any serious analysis and synthesis of Czech. Agreement rules and non-agreement implications for the syntactic system of Czech have so far been considerably exploited in disambiguation of Czech and agreement will be one of the crucial points in partial parsing of Czech based on precise, rule-based morphological disambiguation. The present very short survey tried to show only the basics of a very complex agreement in Czech.