

Logical vs. Natural Language Conjunctions in Czech: A Comparative Study

Katrin Příkrylová, Vladislav Kuboň, and Kateřina Veselovská

Charles University in Prague, Faculty of Mathematics and Physics
Czech Republic

{prikrylova,vk,veselovska}@ufal.mff.cuni.cz

Abstract: This paper studies the relationship between conjunctions in a natural language (Czech) and their logical counterparts. It shows that the process of transformation of a natural language expression into its logical representation is not straightforward. The paper concentrates on the most frequently used logical conjunctions, \wedge and \vee , and it analyzes the natural language phenomena which influence their transformation into logical conjunction and disjunction. The phenomena discussed in the paper are temporal sequence, expressions describing mutual relationship and the consequences of using plural.

1 Introduction and motivation

The endeavor to express natural language sentences in the form of logical expressions is probably as old as logic itself. A very important role in this process is being played by natural language conjunctions and their transformation into logical connectives. The conjunctions are much more ambiguous than logical connectives and thus it is necessary to analyze their role in natural language sentences, in various contexts and types of texts. This paper presents a step towards such analysis for one particular language - Czech.

Let us recall that the fundamental task of logic is to set rules and methods for inferencing and referencing. On the other hand, natural languages serve primarily for communication. Speakers can reach an agreement or understand each other even without a strict adherence to preset rules (regardless whether they are morphological, grammatical or stylistical). A human brain can obtain substantial information also from ill-formed sentences what actually makes them to fulfill their main goal, to serve as a tool for communication. On the other hand, Noam Chomsky introduced in [1] also a famous example *Colourless green ideas sleep furiously* – a grammatically well-formed sentence which does not have any meaning and thus it cannot serve the communication task.

Sentences in any natural language are not isolated, their meaning typically depends on the context in which they appear, on the way how they are pronounced or even on some external factors as, e.g., gestures which accompany it. Natural languages also evolve in time according to the needs of the language community and although each natural language has a set of generally applicable rules (syntactic, stylistic, morphological etc.), there are many ex-

ceptions and irregularities which do not abide the rules as strictly as it is the case in logic.

Primarily due to this difference, the transformation of natural language sentences into their logical representation constitutes a complex issue. As we are going to show in the subsequent sections, there are no simple rules which would allow automation of the process – the majority of problematic cases requires an *individual* approach.

In the following text we are going to restrict our observations to the two most frequently used conjunctions, namely *a* (and) and *nebo* (or).

2 Sentences containing the conjunction *a* (and)

The initial assumption about complex sentences containing the conjunction *a* (and) is inspired by the properties of the corresponding logical connective *and* – we suppose that the two clauses connected by the conjunction express two situations which are valid at the same time. This really holds in a number of complex sentences, as for example here:

Lev je kočkovitá šelma a žije v Africe. (1)
(Lion is a feline and it lives in Africa.)

Jana je ve škole a Honza leží nemocný v posteli. (2)
(Jana is in a school and Honza lies ill in bed.)

In the sentence 1 we have used the so called gnomic present¹. The truth value of the whole sentence is TRUE only in case that both clauses are TRUE, regardless of the context or current situation.

Complex sentences with gnomic present constitute probably the simplest case. It is not necessary to investigate whether the clauses are true or what are the conditions under which they might become true – they are simply true either always or never. Such sentences can be transformed into a logical representation in a simple and straightforward manner.² In the logical representation of the example above we would of course use the construction $A \wedge B$ for the conjunction *a* (and).

¹Present tense can be used also for the so called *extratensal* processes which are valid always, regardless of the current situation. In our example we describe properties of an animal species and its habitat.

²Let us point out that mathematical theorems typically contain gnomic present.

The sentence 2 describes two situations being TRUE exactly in this moment³. The truth value of both sentences can be determined by a *reference* from the language to a real-world, where we will find out whether both clauses describe a valid situation.⁴

None of the two clauses from the sentence 2 is absolutely true (Jana does not spend every minute of her life in the school and Honza is not ill forever). However, when we utter any of these two statements, we do not mean that Jana should stay all the time in the school. The use of the present tense implicitly carries the information that she is there just now, in this moment. If we accept that in order to determine the truth value, we have to look into the real world and also take into the account the time when the sentence was uttered, we could paraphrase the sentence into a more unambiguous variant for example like this:

Jana je právě teď ve škole a Honza nyní leží nemocný v posteli.
(Jana is just now in the school and Honza is now lying ill in bed.) (3)

– i.e. with the added information about time. Such sentence would then correspond to the logical scheme of the conjunction : $A \wedge B$.⁵

Natural languages use, of course, also other tenses – what if we would like to express the same content in the past, for example yesterday?

Jana byla včera ve škole a Honza včera ležel nemocný v posteli.
(Jana was in a school yesterday and Honza was lying ill in bed yesterday.) (4)

On the first sight, there seems to be no substantial problem. The only difference seems to be in the fact that we are not referring to a current moment, but to the moment in the past (in this case, yesterday). However, what if Honza will recover till the next day, will such sentence have the same truth value also tomorrow?

Regardless to what time the expressions refer to, we are interested in them only if they are TRUE in this current moment. We should thus simplify our sentence rather in the following way:

Právě teď platí, že Jana je ve škole, a současně, že Honza leží nemocný v posteli.
(Just now it is true that Jana is in the school, and, at the same time, Honza is lying ill in bed.) (5)

³Of course, only if it is true that Jana is just now in the school and Honza lies ill in his bed.

⁴Determining the truth value of natural language expressions is studied by epistemology, a simple explanation can be found for example in [2].

⁵If we would like to consider tiniest details, we would have to consider also the issue of proper names and singular terms – our sentence does not specify which Jana and Honza we are talking about. More on this topic can be found for example in [3].

Or, we could drop the initial part which we may consider to be implicitly present:

Jana je ve škole, a současně Honza leží nemocný v posteli.
(Jana is in the school, and, at the same time, Honza is lying ill in bed.) (6)

Into such template it is possible to insert also the complex sentence introduced above:

Jana byla včera ve škole a současně Honza včera ležel nemocný v posteli.
(Jana was in a school yesterday and, at the same time, Honza was lying ill in bed yesterday.) (7)

All the complex sentences mentioned above can schematically be described in the form $A \wedge B$. The fact that we can express the mutual relationship of clauses by means of a logical scheme actually means that we can work with them according to logical rules. For example, logical conjunction is commutative – and we really can swap the order of clauses in our complex sentence and still retain the original truth value.

2.1 Violation of a temporal sequence

Unfortunately, the conjunction *a* (and) doesn't appear only in sentences describing actions which are happening in the same moment. All of the following sentences contain *a* (and) as its main conjunction:

Honza spadl a zlomil si ruku.
(Honza fell and broke his arm.) (8)

Jana odemkla a vešla do bytu.
(Jana unlocked and entered the flat.) (9)

Šli jsme na výstavu a potom do kina.
(We have visited an exhibition and then we went to a cinema.) (10)

These sentences apparently aren't commutative. The order of clauses cannot be swapped without affecting the truth value or meaning of the whole sentence. The reason is obvious – both clauses are ordered into a temporal sequence.

The conjunction *a* (and) isn't a logical conjunction in these sentences, although it fulfills one fundamental basic condition – if the whole sentence is supposed to be true, then both clauses also have to be true.

The propositional logic nevertheless cannot cope with sentences of this kind. We might be tempted to attempt to solve this issue by means of the conditional construction *Když... (pak) ...* (When... (then) ...):

Když Jana odemkla, vešla do bytu.
(When Jana unlocked, then she entered the flat.) (11)

and thus to find a certain scheme corresponding to an implication. In natural languages, the modified sentence is equivalent with the original one, but this is true only because the construction *Když...*, (*pak*) ... (When... (then) ...) not necessarily always means an implication. In this particular case, its role is more temporal than conditional.⁶ Transforming the sentence into the scheme $A \rightarrow B$ is thus incorrect. In [4], František Gahér suggests a very simple test whether a particular expression containing the conjunction *a* (and) is a logical conjunction or not. He uses the expression *a současně* (and at the same time).

The sentence:

Jana odemkla a současně vešla do bytu.
(Jana unlocked and at the same time
entered the flat.) (12)

does not make much sense and thus we should not directly transform it into logical conjunction. However, the author itself admits that such simple test is not 100% reliable – the construction:

Gödel se narodil v roce 1906 a zemřel v roce 1978.
(Gödel was born in 1906 and died in 1978.) (13)

actually has all required properties of a conjunction: both clauses must be true if the whole sentence should be true; their order can be changed⁷. However, when we try to replace *a* (and) by the construction *a současně*, (and at the same time), we won't get a meaningful sentence:

*Gödel se narodil v roce 1906 a současně
zemřel v roce 1978.*
(Gödel was born in 1906 and at the same
time died in 1978.) (14)

Let us now return to the original sentence. We have already mentioned that in predicate logic it is impossible to describe it unless we lose an important information about the order of events. What if we would use some other type of logic? The type which seems to be ideally suited for such kind of constructions is the *temporal* logic. It is in fact the propositional logic enriched by the so called *temporal operators*, by means of which we can express a temporal sequence of actions. More information about this kind of logic can be found, e.g., in [5] or [6].

2.2 Disjunction

So far, we have dealt with the conjunction *a* (and) in the cases in which it expressed conjunction. Let us now show that the same natural language conjunction may in some specific cases also serve as a logical disjunction.

⁶The conjunction *když* (when) is then ambiguous.

⁷Although it is more natural to use them in this order. Nevertheless, the variant with the reversed order does not violate neither linguistic rules nor the logical meaning of the sentence.

Let us take the sentence:

*Jestliže Honza neodevzdá diplomovou práci
včas a nepřihlásí se ke státnicím, studia
letos nedokončí.* (15)

(If Honza won't submit the thesis in time
and doesn't subscribe for the state exams,
he won't finish his studies in this year.)

If we would like to preserve the equivalence of *a* (and) and a logical conjunction, we could write this sentence schematically as $(\neg A \wedge \neg B) \rightarrow \neg C$. And indeed, the utterances corresponding to this scheme can be often heard from the Czech native speakers. If we look at the given sentence more closely, we will agree that in order to finish one's studies it is indeed necessary to finish the thesis in time and at the same time to subscribe also for the state exams. If **at least one** of these two conditions is not fulfilled, Honza will not finish his studies. The scheme $(\neg A \wedge \neg B) \rightarrow \neg C$, on the other hand, requires *both* conditions to be invalid in order to obtain FALSE as the truth value of the whole sentence.

It would therefore be more correct to describe the complex sentence schematically as $(\neg A \vee \neg B) \rightarrow \neg C$. The conjunction *a* (and) clearly substitutes logical disjunction in this context. Actually, even in the natural language it would be more correct to use the conjunction *nebo* (or) and to say:

*Jestliže Honza neodevzdá diplomovou práci
včas nebo nepřihlásí se ke státnicím, studia
letos nedokončí.* (16)

(If Honza won't submit the thesis in time
or doesn't subscribe for the state exams,
he won't finish his studies in this year.)

The fact that this error is quite frequent in natural language communication is documented for example in the research of Vlastimil Chytrý [7] conducted among the pupils of basic and secondary schools. Only 11,5 % of them were able to correctly negate the conjunction in the antecedent of the implication, when they were asked to paraphrase it. We can only speculate why the native speakers make this error so often.⁸

2.3 Relation Expression

Let us emphasize that, e.g., a sentence:

Jana a Honza jsou studenti.
(Jana **and** Honza are students.) (17)

is actually a compound sentence:

Jana je studentka a Honza je student.
(Jana is a student and Honza is a student.) (18)

⁸More about the processes in the center of speech of a human brain can be found for example in [8].

i.e. it expresses two utterances.

In the sentence 18, the conjunction *a* (and) is equivalent to a conjunction in logic. However, let us investigate the following examples:

Jana a Honza jsou přátelé.
(Jana and Honza are friends.) (19)

Jana a Honza se milují.
(Jana and Honza love each other.) (20)

Barma a Myanmar je totéž.
(Burma and Myanmar is the same thing)¹⁰ (21)

To rephrase the first sentence as:

Jana je přítelkyně a Honza je přítel.
(Jana is a friend and Honza is a friend.) (22)

makes no sense, since we lose the information about a relationship between Jana and Honza.

In Czech, the second sentence could be rephrased as:

Jana se miluje a Honza se miluje.
(Jana loves herself and Honza loves himself.) (23)

but its meaning is not the same as in case of the original sentence, which is ambiguous in Czech. It contains a reflexive verb *milovat se* (to love someone), which expresses a relationship either between two subjects or of each of them to him/herself particularly. However, we usually use this verb in situations in which we want to express a relationship between two people. Anyway, this example is to illustrate that the same utterance can be formally represented using two different logical schemes. In case we would want to express the second meaning, we would write it down using the means of predicate logic as

$$\text{love_oneself}(Jana) \wedge \text{love_oneself}(Honza)$$

To catch the first meaning, we would have to use not the unary relation, but a binary one

$$\text{Love}(Jana, Honza),$$

which would be in this case symmetrical. Therefore, we would have to abandon the propositional logic to describe this type of sentences.

The last sentence from the list cannot be rephrased as:

Barma je totéž a Myanmar je totéž.
(Burma is the same thing and Myanmar
is the same thing.) (24)

This sentence makes no sense, since the phrase *to be the same thing* again implies a relationship between the entities. These examples actually clearly document the fact that the conjunction *a* (and) used in utterances which express a relationship cannot be used as a conjunction in

logic. The following sentences represent other cases in which the conjunction *a* (and) refers to a relationship between the subjects:

Černé a bílé ponožky se pomíchaly.
(Black and white socks got mixed.) (25)

Jana a Honza se vzali.
(Jana and Honza got married.) (26)

In the above examples we have shown that if the conjunction *a* (and) is used in the utterance which expresses a relationship, it cannot be used as a conjunction in logic.

2.4 Problems with plural

The method of connecting smaller pieces of text than the whole compound sentences which we have introduced above can be called a *distributive* method in the mathematical sense of that term. However, the method is not flawless and we have already shown the examples for which it cannot be applied. We will now demonstrate the imperfections of the method that are not related only to lexicon/semantics (i.e. to particular words which do not let us use the method due to their lexical meaning), but rather to syntax.

Let us consider the following sentence:

Pošt'ák přivezl velký a těžký balík.
(A postman delivered a big and heavy package.) (27)

It is natural to agree with the premise that a postman delivered only one package, which was big and heavy at the same time.¹¹ However, if we divide the sentence into two propositions:

Pošt'ák přivezl velký balík a pošt'ák přivezl těžký balík.
(A postman delivered a big package and
a postman delivered a heavy package)
(28)

the most natural interpretation would probably be that a postman delivered two packages, one of them big and the other one heavy.

The distinction is even more obvious in the following sentence:

Na ulici stálo modré a zelené auto.
(There was a blue and green car parking
on the street.) (29)

Although the word *car* is used here in singular, we would probably say that there were two cars parking on the street, one of them blue and the other one green. In case the author would use plural:

Na ulici stála modrá a zelená auta.
(There were blue and green cars parking
on the street.) (30)

¹¹That would probably be the first interpretation which would come to our mind without thinking about any further meanings.

we would probably come into conclusion that there were even more than two cars parking on the street.

These examples demonstrate that when connecting two adjectives, the interpretation of the conjunction *a* (and) is not clear. Whereas in the first example it is a description of one object having two characteristics, in the second example we describe two different objects having two different characteristics. However, we assign the same activity (same predicate) to both of these objects. The type of the structure is given by the particular adjectives. It is not common in a real word that the car would be both blue and green at the same time.¹² In the case when it would be a dirty and scratched car, it would probably be perceived as only one vehicle.

More syntactic problems are connected with a plural. While in the sentence:

Na ulici stálo špinavé a poškrábané auto.
(There was a dirty and scratched car parking
on the street.) (31)

we assign both characteristics to only one object (a car), in the sentence:

Na ulici stála špinavá a poškrábaná auta.
(There were dirty and scratched cars parking
on the street.) (32)

we do not insist on assigning both characteristics to all of the vehicles.

The second sentence thus cannot be interpreted as a conjunction, but rather as a disjunction.¹³ In the case of a plural we cannot consider this feature as a specific property of particular adjectives. This phenomena is not related to a specific semantics of given lexemes, but concerns all the adjectives in their fullness.

Below we list some other sentences which should be considered:

Článek se zabývá aktivními a pasivními příjmy.
(The article discusses the active and passive
incomes.) (33)

Sešli se tam všichni místní slavní a bohatí lidé.
(All the local famous and rich people met up
at the event.) (34)

¹²In this case, the car would probably rather be described as *blue-green*.

¹³However, we still need to take into account the level of which we speak. Whereas in connection with the noun (or a noun phrase) the adjective is attributed to we talk about disjunction (it is not required for both objects to have both characteristics), in context of the whole sentence the conjunction *a* (and) behaves as a conjunction again. It means that if there would be only dirty cars parking on the street, we would be wondering where are the scratched ones mentioned in the sentence as well.

*Jako cestovatel se dostal na mnohem zajímavější
a podivuhodnější místa.*
(As a traveler, he got to far more
interesting and remarkable places.) (35)

3 Interpretation of the Sentences Containing the Conjunction *nebo* (or)

As we have shown in the previous section, the interpretation of the conjunction *a* (and) is not an easy task. Surprisingly, the conjunction *nebo* (or) behaves more systematically.

The conjunction *nebo* (nebo) can be interpreted in two ways:

- as a disjunction,
- as an exclusive disjunction.

Apart from English, Czech language has a rather strict rules distinguishing between these two cases.¹⁴ If there is *nebo* (or) following the comma, it is an exclusive disjunction. In all the other cases, it is considered a common disjunction:

Čertovi se také říká d'ábel nebo satan.
(The demon is also called a devil or a Satan.) (36)
– a disjunction

Honza přijede ve středu, nebo až ve čtvrtek.
(Honza is coming on Wednesday,
or on Thursday.) – exclusive disjunction (37)

Naturally, in the spoken language we do not have a chance to find out whether there is a comma in the sentence or not.¹⁵ Therefore, we have a lexical distinction at our disposal: the exclusive *nebo* (or) becomes a correlative conjunction, namely *bud'–nebo* (either–or):

Honza přijede bud' ve středu, nebo až ve čtvrtek.
(Honza is coming either on Wednesday
or on Thursday.) – exclusive disjunction (38)

Conjunction *nebo* (or) can also be a part of the more complex connection which can be further expressed using other logical conjunction. For illustration, see the following sentence:

At' už Honza přijede, nebo ne, oslava se bude konat.
(Whether Honza is coming or not, we will
throw the party.) (39)

¹⁴In English, we use a comma preceding the conjunction *or* when it connects two independent sentences, regardless the relationship between them.

¹⁵In Czech, we place commas based on structural rules, i.e. not in places where there is a natural break in spoken utterance.

*At' už Honza přijede ve středu, nebo ve čtvrtek,
rozhodně navštíví také prarodiče.* (40)
(Whether coming on Wednesday or Thursday,
Honza will definitely drop by his grandparents.)

As for the sentence 39, we can write down the proposition using a special logical conjunction **Maybe and** (MA, truth depends on second proposition) described for example in [9]: *Honza is coming MA we will throw a party.*

The sentence 40 is however much more complex. Although it expresses the contrast of the two possibilities, one of them is not a negation of another. Therefore we cannot use the conjunction MA, since it is a binary conjunction and we need to connect three propositions.¹⁶ The second sentence can be transformed into a logical notation in the following way:

*(Honza přijede ve středu \oplus Honza přijede ve čtvrtek) \wedge
Honza rozhodně navštíví prarodiče. ((Honza is coming on
Wednesday \oplus Honza is coming on Thursday) \wedge Honza
will drop by his grandparents.)* (41)

Therefore, we would have to use the exclusive disjunction again.¹⁷

4 Conclusion

In this article, we have discussed the interpretation of natural language sentences using the means of logic. We have shown that although some of the logical conjunction names are motivated by the natural language conjunctions and they quite often have similar meaning, it is not possible to translate them from natural language to logic directly. Especially for the conjunction *a* (and) we have introduced more complex problems (i.e. the issue of relations, plurals or sequence of tenses) which prevent us from identifying *a* (and) with a logical conjunction.

Also, we have brought an important analysis of the possibilities (and problems) which have to be considered when working beyond sentential level. We have shown how to transform these structures so that they could be described using the means of the propositional logic (which takes only the propositions – or, in other words, sentences).

¹⁶1. Honza is coming on Wednesday. 2. Honza is coming on Thursday. 3. Honza will drop by his grandparents.

¹⁷Conjunction MA can be also expressed using a set of conjunctions $\{\wedge, \vee, \neg\}$. It is also interesting to consider whether the construction *At' už (...), nebo (...)* (Whether (...) or (...)) can be captured using a common disjunction or using the exclusive one. In case we are describing an indisputable system (such as propositional logic), we already know that the situation $A \wedge \neg A$ is impossible, so both versions of the translation – with \vee and with \oplus – are equivalent if there is the same formula in the connection (or, more precisely, the formula and its negation). Finally, we have to mention that there were both variants (with and without a comma) found in the corpus.

5 Acknowledgments

This research was supported by the grant GA15-06894S of the Grant Agency of the Czech Republic and by the SVV project number 260 224. This work has been using language resources stored and/or distributed by the LINDAT-Clarín project of MŠMT (project LM2010013).

References

- [1] Chomsky, N.: Syntactic Structures. Werner Hildebrandt, Berlin (2002)
- [2] Dummett, M.: Origins of Analytical Philosophy. Harvard University Press, Cambridge, Massachusetts (1996)
- [3] Marvan, T.: Otázka významu. Togga, Praha (2010) In Czech.
- [4] Gahér, F.: Logika pre každého. Iris, Braatislava (1995) in Slovak.
- [5] Øhrstrom, P., Hasle, P.F.V.: Temporal Logic: From Ancient Ideas to Artificial Intelligence. Kluwer Academic Publishers, Dordrecht, Netherlands (1995)
- [6] Sag, I., Wiesler, S.: Temporal connectives and logical form. In: Proceedings of the Fifth Annual Meeting of the Berkeley Linguistics Society, Berkeley (1979) 336–349
- [7] Chytrý, V.: Logika, hry a myšlení. Univerzita J. E. Purkyně, Ústí nad Labem (2015) in Czech.
- [8] C.L.Baker, McCarthy, J.J.: The Logical Problem of Language Acquisition. The MIT Press, Cambridge, Massachusetts (1981)
- [9] van Wijk, M.: Logical connectives in natural language. Doctoraalscriptie Algemene Taalwetenschap. Universiteit Leiden Faculteit der Letteren, Leiden (2006) Available online: <https://www.era.lib.ed.ac.uk/bitstream/handle/1842/5822/VanWijk2006.pdf>.