

# Comprehensibility limits on centre-embedded structures

Bogdan Babych <[smlbb@leeds.ac.uk](mailto:smlbb@leeds.ac.uk)>

**Summary.** The issue of centre embedding (in particular, self-embedding, e.g., in the sentence: *The rat [the cat bit] ate the cheese*) is central to the problem of finding a minimal necessary complexity for a model of natural language. It was proved that only unlimited self-embedding causes language to be context-free. Less complex finite-state models of natural language could be considered inappropriate only if there is a task to accommodate unlimited self-embedding.

However, multiple self-embedding often becomes incomprehensible for human recipients (e.g., *\*The rat [<sub>1</sub> the cat [<sub>2</sub> the dog chased <sub>2</sub>] bit <sub>1</sub>] ate the cheese*), but not always. (E.g., the sentence: *What [<sub>1</sub> the woman [<sub>2</sub> that John married <sub>2</sub>] likes <sub>1</sub>] is smoked salmon*. – is comprehensible, though it also has 2 levels of centre embedding). In our talk a linguistic explanation is proposed which accounts for failures in comprehension of some classical examples of centre-embedded relative clauses. In our model such failures are related to impossibility of resolving co-reference of pronouns and noun phrases in clauses on different levels; the sentences in these cases are considered ungrammatical.

Higher levels of centre embedding are comprehensible in phrases where parsing is not disrupted by possible failures in reference resolution. To find out comprehensibility limits of self-embedded structures, we investigate artificially generated Ukrainian noun phrases. Centre embedding of such phrases is possible and grammatical, since Ukrainian, as many other Slavic languages, allows a specific kind of word order variation: permutation of branches in same sub-trees in the syntactic structure).

Real comprehensibility limits of centre embedding are shown to conform to G.Miller's model of short-term memory (that can hold up to  $7 \pm 2$  units), so 3 or sometimes 4 levels of centre embedding still can be comprehensible. Levels of centre embedding higher than 4 cannot take part in human communication, since they are too difficult for generation and understanding. These results could be used as guidelines for creating an appropriate finite-state model of natural language, and also for developing a parsing strategy, which should account for failures in reference resolution in self-embedded relative clauses.

## 1. Acceptable centre-embedded sentences:

(1.1) *The dog [that the girl scared] ran away*<sup>1</sup>.

(1.2) *The bird [that the mouse chased] scared the cat.*

Centre-embedded sentences can also have no overt conjunctions:

(2) *The rat the cat bit ate the cheese*

## Unacceptable centre-embedded sentences:

(3.1) *\*? The cat [<sub>1</sub> that the bird [<sub>2</sub> that the mouse chased <sub>2</sub>] scared <sub>1</sub>] ran away*

(3.2) *\*? The rat the cat the dog chased bit ate the cheese.*

There are no problems with understanding right-branching structures:

(4.1) *The mouse chased the bird [that scared the cat [that ran away]];*

(4.2) *The cat chased the mouse [that scared the cat [that saw the dog [that ate the pumpkin]]] ...*

---

## 2. The first explanation

[Miller, G.A. and Chomsky, N., 1963. Introduction to the formal analysis of natural languages. In: Luce, D.R., Bush R.R., Galanter, E., editors, *Handbook of Mathematical Psychology*, volume II, John Wiley, New York]

---

<sup>1</sup> English examples and the overview taken from [Lewis, R.L. 1996, A theory of grammatical but unacceptable embedding. URL: "[citeseer.nj.nec.com/richard96theory.html](http://citeseer.nj.nec.com/richard96theory.html)"]

1. On the level of **linguistic competence** (knowledge of language), multiple centre embedding is allowed.

On the level of **linguistic performance** (real processes of speaking / hearing), human short-term memory is not sufficient to process multiple centre-embedded structures

2. Sentence (3) is **grammatical**, but not acceptable from the **performance** point of view.

3. Sentence (3) is different from ungrammatical sentences, e.g. (5), (6):

(5) \* *The dogs runs away.*

(6) \* *I am eager John to be here.*

vs. grammatical sentences (7), (8), (9):

(7) *I believe John to be here.*

(8) *I am eager for John to be here.*

(9) *I am eager to be here.*

---

### 3. Why this problem is important:

The solution could provide an answer to the questions:

**Question 1:** What is the complexity of the appropriate model of natural language?

- finite-state,
- context-free,
- mildly context sensitive, e.g., TAGs
- context sensitive

Finite-state models of the natural language are not appropriate **only if** we are required to accommodate unlimited centre embedding (self-embedding in particular)<sup>2</sup>.

In language material (real texts, e.g. corpora) unlimited centre embedding could not be found.

However, do we need separate models for linguistic competence (e.g., which allows unlimited centre embedding) and linguistic performance (which introduces psychological limits on processing centre-embedded structures)? Since unlimited centre embedding cannot be neither produced nor understood, it is not evident that linguistic competence is a separate mechanism from the general cognitive mechanisms. It is possible that a finite-state model of natural language, which accommodates highest comprehensible levels of centre embedding, is psychologically appropriate:

**Question 2:** Is it possible to learn the grammar of the right complexity only from the language material (oral and written texts), using general learning techniques?

- are we born with a kind of a "universal grammar" or a language instinct, and use it while learning language, or
- do we use general learning techniques and cognitive abilities to learn language, in a similar way as we learn other skills?

It is much easier to learn finite-state models from language material with existing machine-learning techniques.

Language material does not contain unlimited levels of centre embedding (is not clear what is a psychologically appropriate model for learning it).

-- Finding out how the comprehensibility of centre embedding is limited will allow establishing appropriate learning model, without a need for redundant computational complexity.

---

### 4. Counterexamples of acceptable instances of multiple centre-embedding, which undermine the previous explanation<sup>3</sup>:

---

<sup>2</sup> Proved in [Chomsky, N., 1959, On certain formal properties of grammars. In: Information and control, 2:137-167]

<sup>3</sup> [Cowper, E.A., 1976. Constraints on sentence complexity: A model for syntactic processing. PhD thesis. Brown University]; [Gibson, E.A., 1991. A computational theory of human linguistic processing: Memory limitations and

- (10) *That* [<sub>1</sub> *the food* [<sub>2</sub> *that John ordered* <sub>2</sub>] *tasted good* <sub>1</sub>] *pleased him*.  
 (11) *What* [<sub>1</sub> *the woman* [<sub>2</sub> *that John married* <sub>2</sub>] *likes* <sub>1</sub>] *is smoked salmon*.  
 (12) *Say, do you think that*  
 [<sub>0</sub> *the report that* [<sub>1</sub> *the stuff* [<sub>2</sub> *they put in diet soda* <sub>2</sub>] *causes cancer* <sub>1</sub>] *is a hoax* <sub>0</sub>]?

## 5. Other explanations [Lewis, R.L. 1996]:

### Linguistic metrics

Theories, which define metrics over syntactic structure and predict the perceived relative complexity of parsing different structures, e.g.:

- degree of self-embedding (Miller, G., Chomsky, N., 1963)
- ratio of nodes to terminals,
- nodes open to further expansion,
- number of NPs playing two different grammatical roles

### Architectures

Theories, which define conditions for acceptable and difficult embedding in terms of some specific computational architecture for parsing, with some limited resource, e.g.:

- push-down automaton (stack cells are limited to 7) (Yngve, 1960, Miller, 1956 models<sup>4</sup>)
- subroutine architecture (return address memory is limited)

### Parsing strategies

Theories, which posit specific strategy for uncovering grammatical relations and show how the strategy breaks down in cases of deep centre embedding, e.g.:

- interpreting a series of NPs as a conjoined series of nouns rather than as the subjects of independent clauses

(We present a linguistic explanation, which could be developed into a parsing strategy explanation)

## 6. Ukrainian word order: possibility of generating and testing centre-embedded structures of different types. (Many other Slavic languages also have these types of word order).

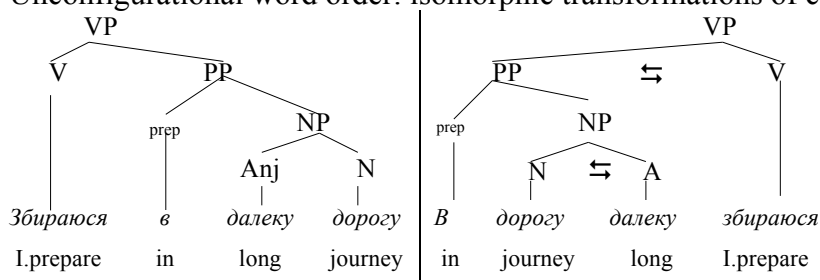
Direct word order:

(1) *Збираюся в далеку дорогу* (V [prep Adj N]<sub>obj</sub>)

I.prepare<sub>pres.sing.pl</sub> [into<sub>prep/acc</sub> long<sub>Adj.acc.sing.fem</sub> journey<sub>Noun.acc.sing.fem</sub>]

I prepare for a long journey

Unconfigurational word order: isomorphic transformations of constituency trees, e.g.:



Nearly all possible isomorphic transformations are allowed; permutations of constituents within phrases changes the *direct* word order into an *unconfigurational* word order:

(2) *В далеку дорогу збираюся* ([prep Adj N]<sub>obj</sub> ⇌ V)

[into<sub>prep/acc</sub> long<sub>Adj.acc.sing.fem</sub> journey<sub>Noun.acc.sing.fem</sub>] ⇌ I.prepare<sub>pres.sing.1</sub>

(a verb follows an object on the sentence level)

processing breakdown. PhD thesis. Carnegie Mellon. Available as Centre for Machine Translation technical report CMU-CMT-91-125]; [Labov, W., 1973. The place of linguistic research in American society. In Hamp, E.P., editor, Themes in Linguistics: The 1970s. Mouton, The Hague]

<sup>4</sup> [Miller, G.A., 1956, The magical number seven plus minus two: Some limits on our capacity for processing information. In: The psychological review, 63(2):81-97]; [Yngve, V.H., 1960, A model and a hypothesis for language structure. In: Proceedings of the American Philosophical Society, 104(5):444-466].

(3) Збираюся в дорогу далеко ( V [prep [N ⇔ Adj]])

I.prepare<sub>pres.sing.pl</sub> [into<sub>prep/acc</sub> journey<sub>Noun.acc.sing.fem</sub> ⇔ long<sub>Adj.acc.sing.fem</sub>]

(the adjective follows a noun; the predicative meaning of the adjective is increased in this case)

(4) В дорогу далеко збираюся ([prep [N ⇔ Adj]] ⇔ V)

[into<sub>prep/acc</sub> journey<sub>Noun.acc.sing.fem</sub> ⇔ long<sub>Adj.acc.sing.fem</sub>] ⇔ I.prepare<sub>pres.sing.pl</sub>

(permutations on different levels of the sentence structure)

## 7. Generating centre-embedded noun phrases extended with prepositional phrases:

Embedded constructs have the direct word order; synonymous constructs with right branching have unconfigurational word order.

### Centre embedding of level (1)

(5a) NP<sub>[AP<sub>[A відомий PP<sub>[у західних країнах]]</sub> N письменник]</sub> --</sub>

known<sub>A.NOM.SING</sub> in<sub>PREP</sub> western<sub>A.LOC.PLUR</sub> countries<sub>N.LOC.PLUR</sub> writer<sub>N.NOM.SING</sub>

⇔ (5b) NP<sub>[N письменник, ⇔ AP<sub>[A відомий PP<sub>[у західних країнах]]</sub>]]</sub></sub>

writer<sub>N.NOM.SING</sub> known<sub>A.NOM.SING</sub> in<sub>PREP</sub> western<sub>A.LOC.PLUR</sub> countries<sub>N.LOC.PLUR</sub>

‘A writer, known in western countries’

### 8. Centre embedding of level (2)

(6) (a) NP<sub>[6AP<sub>[5 отримана</sub></sub>

PP<sub>[4 від</sub> NP<sub>[3 AP<sub>[2 відомого</sub></sub>

PP<sub>[1 у західних країнах<sub>1</sub>]]<sub>2</sub></sub>

письменника<sub>3</sub>]]<sub>4</sub>]]<sub>5</sub>

книга<sub>6</sub>]

received<sub>A.NOM.SING.FEM</sub> from<sub>PREP</sub> known<sub>A.NOM.GEN.MASC</sub> in<sub>PREP</sub> western<sub>A.LOC.PLUR.FEM</sub> countries<sub>N.LOC.PLUR.FEM</sub>

writer<sub>N.GEN.SING.MASC</sub> book<sub>N.NOM.SING.FEM</sub>

⇔ (b) NP<sub>[N книга,</sub>

↓↑

AP<sub>[отримана від</sub> NP<sub>[N письменника,</sub>

↓↑

AP<sub>[відомого у західних країнах]]]]</sub>

book<sub>N.NOM.SING.FEM</sub> received<sub>A.NOM.SING.FEM</sub> from<sub>PREP</sub> writer<sub>N.GEN.SING.MASC</sub> known<sub>A.NOM.GEN.MASC</sub> in<sub>PREP</sub>

western<sub>A.LOC.PLUR.FEM</sub> countries<sub>N.LOC.PLUR.FEM</sub>

‘A book, which was received from the writer, who is known in western countries’

### 9. Centre embedding of level (3)

(7) (a) NP<sub>[9AP<sub>[8 характеризований</sub></sub>

PP<sub>[7 в</sub> NP<sub>[6AP<sub>[5 отриманій</sub></sub>

PP<sub>[4 від</sub> NP<sub>[3 AP<sub>[2 відомого</sub></sub>

PP<sub>[1 у західних країнах<sub>1</sub>]]<sub>2</sub></sub>

письменника<sub>3</sub>]]<sub>4</sub>]]<sub>5</sub>

книзі<sub>6</sub>]]<sub>7</sub>

персонаж<sub>8</sub>]]<sub>9</sub>

characterized<sub>A.NOM.SING.MASC</sub> in<sub>PREP</sub> received<sub>A.LOC.SING.FEM</sub> from<sub>PREP</sub> known<sub>A.NOM.GEN.MASC</sub> in<sub>PREP</sub> western<sub>A.LOC.PLUR.FEM</sub>

countries<sub>N.LOC.PLUR.FEM</sub> writer<sub>N.GEN.SING.MASC</sub> book<sub>N.LOC.SING.FEM</sub> personage<sub>N.NOM.SING.MASC</sub>

⇔ (b) NP<sub>[N персонаж,</sub>

↓↑

AP<sub>[характеризований PP<sub>[в</sub> NP<sub>[N книзі,</sub></sub>

↓↑

AP<sub>[отриманій від</sub> NP<sub>[N письменника,</sub>

↓↑

AP<sub>[відомого у західних країнах]]</sub>

personage<sub>N.NOM.SING.MASC</sub> characterized<sub>A.NOM.SING.MASC</sub> in<sub>PREP</sub> book<sub>N.LOC.SING.FEM</sub> received<sub>A.LOC.SING.FEM</sub> from<sub>PREP</sub>

writer<sub>N.GEN.SING.MASC</sub> known<sub>A.NOM.GEN.MASC</sub> in<sub>PREP</sub> western<sub>A.LOC.PLUR.FEM</sub> countries<sub>N.LOC.PLUR.FEM</sub>

‘A personage, who is characterized in a book, which was received from the writer, who is known in western countries’

## 10. Centre embedding of level (4)

(8) (a)  $NP[12AP[11пов'язана$

$PP[103NP[9AP[8$  *характеризованим*  
 $PP[76$   $NP[6AP[5$  *отриманій*  
 $PP[4від$   $NP[3$   $AP[2відомого$   
 $PP[1у$  *західних країнах* $1]2]$   
*письменника* $3]4]5]$   
*книзі* $6]7]$   
*персонажем* $8]9]10]11]$   
*проблематика* $12]$

related<sub>A.NOM.SING.FEM</sub> to<sub>PREP</sub> characterized<sub>A.INS.SING.MASC</sub> in<sub>PREP</sub> received<sub>A.LOC.SING.FEM</sub> from<sub>PREP</sub> known<sub>A.NOM.GEN.MASC</sub>  
in<sub>PREP</sub> western<sub>A.LOC.PLUR.FEM</sub> countries<sub>N.LOC.PLUR.FEM</sub> writer<sub>N.GEN.SING.MASC</sub> book<sub>N.LOC.SING.FEM</sub> personage<sub>N.INS.SING.MASC</sub>  
problems<sub>N.NOM.SING.FEM</sub>

⇒ (b)  $NP[N$  *проблематика,*

↓↑  
 $AP[пов'язана з$   $NP[N$  *персонажем,*  
↓↑  
 $AP[характеризованим в$   $NP[N$  *книзі,*  
↓↑  
 $AP[отриманій від$   $NP[N$  *письменника,*  
↓↑  
 $AP[відомого у західних країнах]]]]]]]]]$

problems<sub>N.NOM.SING.FEM</sub> related<sub>A.NOM.SING.FEM</sub> to<sub>PREP</sub> personage<sub>N.INS.SING.MASC</sub> characterized<sub>A.INS.SING.MASC</sub> in<sub>PREP</sub>  
book<sub>N.LOC.SING.FEM</sub> received<sub>A.LOC.SING.FEM</sub> from<sub>PREP</sub> writer<sub>N.GEN.SING.MASC</sub> known<sub>A.NOM.GEN.MASC</sub> in<sub>PREP</sub>  
western<sub>A.LOC.PLUR.FEM</sub> countries<sub>N.LOC.PLUR.FEM</sub>

'Problems, that are related to a personage, who is characterized in a book, which was received from the writer, who is known in western countries'

## 11. Centre embedding of level (5)

(9) (a)  $NP[15AP[14представлена$

$PP[13у$   $NP[12AP[11пов'язаній$   
 $PP[103NP[9AP[8$  *характеризованим*  
 $PP[76$   $NP[6AP[5$  *отриманій*  
 $PP[4від$   $NP[3$   $AP[2відомого$   
 $PP[1у$  *західних країнах* $1]2]$   
*письменника* $3]4]5]$   
*книзі* $6]7]$   
*персонажем* $8]9]10]11]$   
*проблематиці* $12]13]14]$   
*концепція* $15]$

represented<sub>A.NOM.SING.FEM</sub> in<sub>PREP</sub> related<sub>A.LOC.SING.FEM</sub> to<sub>PREP</sub> characterized<sub>A.INS.SING.MASC</sub> in<sub>PREP</sub> received<sub>A.LOC.SING.FEM</sub>  
from<sub>PREP</sub> known<sub>A.NOM.GEN.MASC</sub> in<sub>PREP</sub> western<sub>A.LOC.PLUR.FEM</sub> countries<sub>N.LOC.PLUR.FEM</sub> writer<sub>N.GEN.SING.MASC</sub>  
book<sub>N.LOC.SING.FEM</sub> personage<sub>N.INS.SING.MASC</sub> problems<sub>N.LOC.SING.FEM</sub> conception<sub>N.NOM.SING.FEM</sub>

(b)  $NP[N$  *концепція,*

↓↑  
 $AP[представлена у$   $NP[N$  *проблематиці,*  
↓↑  
 $AP[пов'язаній з$   $NP[N$  *персонажем,*  
↓↑  
 $AP[характеризованим в$   $NP[N$  *книзі,*  
↓↑  
 $AP[отриманій від$   $NP[N$  *письменника,*  
↓↑  
 $AP[відомого у західних країнах]]]]]]]]]$

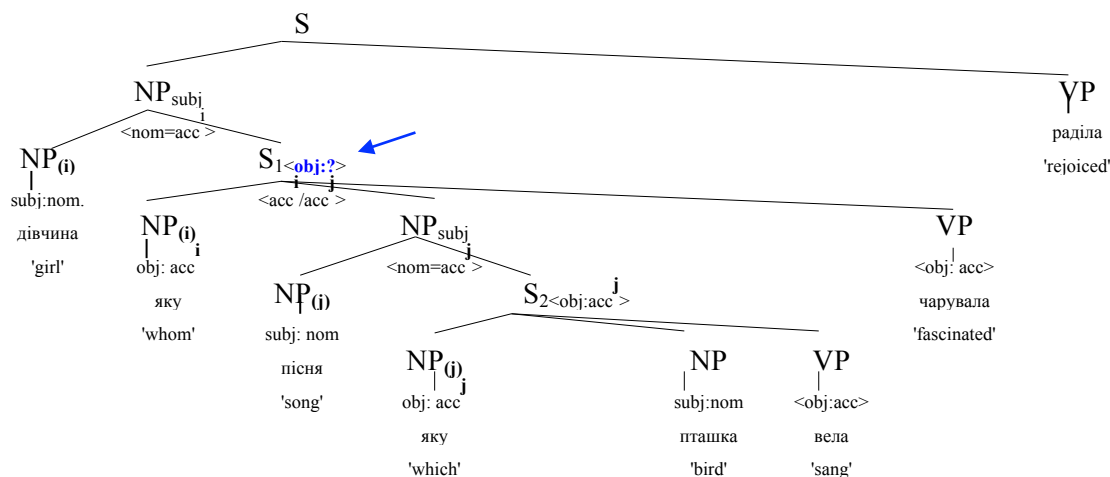
conception<sub>N.NOM.SING.FEM</sub> represented<sub>A.NOM.SING.FEM</sub> in<sub>PREP</sub> problems<sub>N.LOC.SING.FEM</sub> related<sub>A.LOC.SING.FEM</sub> to<sub>PREP</sub>  
personage<sub>N.INS.SING.MASC</sub> characterized<sub>A.INS.SING.MASC</sub> in<sub>PREP</sub> book<sub>N.LOC.SING.FEM</sub> received<sub>A.LOC.SING.FEM</sub> from<sub>PREP</sub>  
writer<sub>N.GEN.SING.MASC</sub> known<sub>A.NOM.GEN.MASC</sub> in<sub>PREP</sub> western<sub>A.LOC.PLUR.FEM</sub> countries<sub>N.LOC.PLUR.FEM</sub>



## 16. An unacceptable sentence, centre-embedding of level (2):

(11) \* [<sub>S</sub> [<sub>NP:subj</sub> Дівчина,  
[<sub>S</sub> яку [<sub>NP:subj</sub> пісня,  
[<sub>S</sub> яку пташка вела]],  
V:pred чарувала]],  
V:pred раділа].

girl<sub>N.NOM.SING.FEM</sub> whom<sub>PRON.ACC.SING.FEM</sub> song<sub>N.NOM.SING.FEM</sub> which<sub>PRON.ACC.SING.FEM</sub> bird<sub>N.NOM.SING.FEM</sub>  
sang<sub>V.SING.FEM.PAST.IMPERF</sub> fascinated<sub>V.SING.FEM.PAST.IMPERF</sub> rejoiced<sub>V.SING.FEM.PAST.IMPERF</sub>  
'A girl, who was fascinated by the song sung by the bird, was happy.'



On the level of the second embedded subordinate clause  $S_2$  the identification of the accusative object " $\text{obj:acc}^j$ " is successful, like in (10). But we suggest that on the level of the first embedded clause  $S_1$  the accusative object required by the verb "чарувала" ('fascinated') cannot be identified, because **the "accusative" case feature is already present in both sister nodes of the verb**: in the subject  $\text{NP}_{\text{subj}} \langle \text{nom=acc}^j \rangle$  and in the pronoun object  $\text{NP}_{(i)} \text{obj:acc}^i$  яку 'whom'. As a result, the recipient cannot answer the question 'who was fascinated?' – 'the song' or 'the girl' because of inconsistency in feature structures on the top level of the first embedded subordinate clause:  $S_1 \langle \text{obj:??} \rangle$ . This explains why the two levels of embedding of subordinate clauses become incomprehensible (particularly – for case when the clauses modify subjects in a higher level clause, and this subject is co-referential with an object pronoun inside such subordinate clause).

## 17. Predictions made by this model: unacceptable centre embedding of level (1):

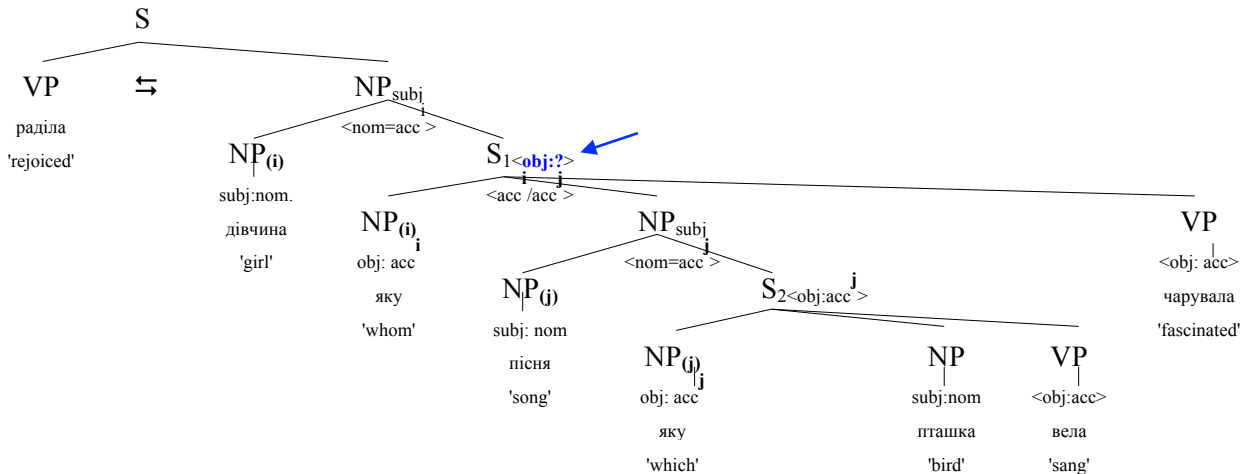
Comprehensiveness of such structures should not depend on the level of centre embedding (the level (1) should also be unacceptable). So we could swap the order of branches in the top-level clause, moving the verb to the front, and the sentence should remain incomprehensible

In other words, we can change the order of the subject and verb in the main clause: "Subj. V" into "V  $\rightleftharpoons$  Subj" (as it is allowed by isomorphic unconfigurational transformation of 11), so the level of embedding will be 1, but the sentence will remain incomprehensible.

This prediction is correct. Indeed, the sentence (12) – with 2 relative clauses and only 1 level of embedding – remains incomprehensible:

(12) \* [<sub>S</sub> Раділа, V:pred  $\rightleftharpoons$  дівчина, N:subj, [<sub>S</sub> яку Pron:obj пісня, N:subj, [<sub>S</sub> яку пташка вела], чарувала V:pred]].

rejoiced<sub>V.SING.FEM.PAST.IMPERF</sub> girl<sub>N.NOM.SING.FEM</sub> whom<sub>PRON.ACC.SING.FEM</sub> song<sub>N.NOM.SING.FEM</sub> which<sub>PRON.ACC.SING.FEM</sub>  
bird<sub>N.NOM.SING.FEM</sub> sang<sub>V.SING.FEM.PAST.IMPERF</sub> fascinated<sub>V.SING.FEM.PAST.IMPERF</sub>



(What is important here is the presence of the co-referential accusative object in the first subordinate clause and subject in the main clause).

The English sentence (13) (that is derived from 3.2) with the phrase structure similar to (12) and with only 1 level of embedding is also incomprehensible (?):

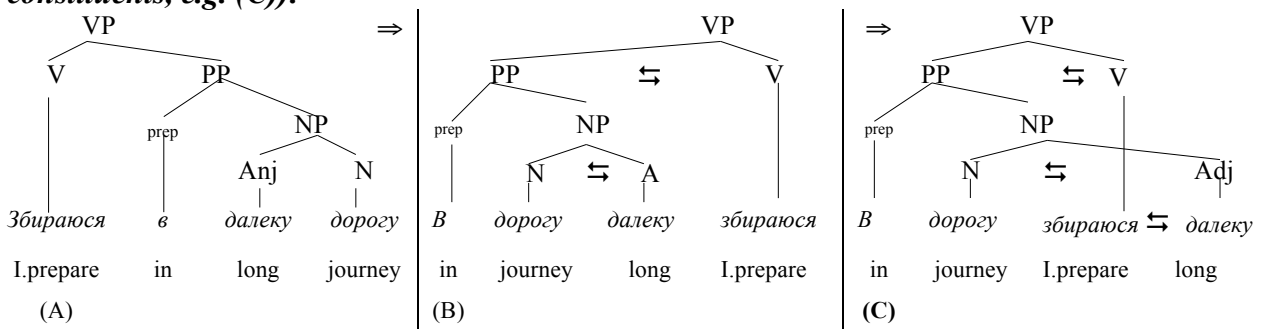
(13) \* *The cheese was eaten by the rat [the cat [the dog chased] bit].*

This proves that comprehensiveness of this type of sentences depends on the feature coherence in the phrase structure and is not directly related to the level of embedding and the recipient's short-term memory size.

## 18. Conclusions

- Finite state models could account for the complexity of natural language syntax, if they accommodate up to 4 levels of centre embedding; higher levels of embedding become incomprehensible and cannot take part in human communication.
- Reference resolution is crucial for comprehensibility of relative clauses with centre embedding: reference resolution becomes impossible during processing of centre embedded constructs of certain types, e.g., subject relative clauses connected by (overt or covert) pronouns in the objective case. Sentences like (3), (12), (13) could be qualified as ungrammatical; they become incomprehensible because of failed reference resolution. Other types of centre-embedded constructs tolerate up to 4 levels of centre embedding.
- A parsing strategy that models human understanding has to take into account the reference resolution issues during sentence processing.

## 19. Further research: Syntactic complexity of unprojective structures (that have discontinuous constituents, e.g. (C)):



Unprojective sentences in modern Ukrainian are stylistically distinct from sentences with the two other types of word order. Native speakers feel stylistic ‘unrestrictedness’ of sentences with the unprojective word order, the ‘peripheral’ character of these sentences, which nevertheless remain grammatical and are widely accepted in ‘unrestricted’ styles of Ukrainian, such as poetry. E.g. in the line taken from the poem of the modern Ukrainian poet Vasyl' Stus, presented in (14), the noun phrase *злупа ніч* ‘the late night’ becomes discontinuous because its constituents are separated by the verb:

(14) *I [злупа.. облягає... ніч] пекельний край і крик пекельний*  
 and late surrounds night infernal land and cry infernal  
 A.NOM. V.3<sup>rd</sup>.SING. N.NOM. A.ACC.SING. N.ACC. N.ACC. A.ACC.



SING.FEM      PRES      SING.FEM      MASC      SING.MASC      SING.MASC      SING.MASC

‘And the late night surrounds the infernal land and the infernal cry’

In modern Ukrainian unprojective sentences are used mainly in poetry, in spontaneous speech, and in the direct speech of characters in prosaic texts. However, in Old Ukrainian of 17<sup>th</sup> century, during a baroque period, it was a sign of a formal "high" written style. Probably, this feature can be attributed to a baroque "way of thinking", which was characterised by "joining of what could not be joined". E.g., the sentence (15) opens a chronicle written by a Cossack historian Samiylo Velychko about the events of the Ukrainian war for independence of 1648-1656 (led by Bohdan Khmelnytsky):

(15)

Нум	зачну	NP[ <sub>i</sub> наступуюче ...	NP[ <sub>j</sub> воєнних...	Хмельницького	...діл <sub>i</sub> ]	з	Поляки	...поведеніє <sub>i</sub> ]
by	I.will	following	military	of	actions	with	Poles	conduct
this	start	A.ACC.SING.NEUT	A.GEN.PLUR	Khmelnysky	N.GEN.		N.INS.	N.ACC
N.INS.	V.SING.			N.GEN.SING	PLUR		PLUR	SING.NEUT
SING	1st.FUT							

‘By this I will start [to relate] the following conduct of military actions by Khmelnytsky against Poles’.

The phrase "*the following conduct of military actions by Khmelnytsky*" has an unprojective word order with literal translation: "*following of military by Khmelnytsky actions conduct*" (a PP "*with Poles*" is a complement of the word "*actions*", like the genitive "*of Khmelnytsky*"). Morphological and semantic features allow the recipient finding appropriate syntactic links and reconstructing the direct word order with the corresponding isomorphic constituent structure, as in (16):

(16) Нум зачну NP[<sub>i</sub> наступуюче поведеніє <sub>i</sub>] NP[<sub>j</sub> воєнних діл <sub>i</sub>] Хмельницького з Поляки.

With.this<sub>pron.ins.sing.masc</sub> I.will.start<sub>v.sing.1st.fut</sub> following<sub>Adj.acc.sing.neut</sub> conduct<sub>N.acc.sing.neut</sub> military<sub>Adj.gen.plur</sub> actions<sub>N.gen.plur</sub>.  
 Khmelnytsky<sub>N.gen.sing.mask</sub> with<sub>Prep.inst</sub> Poles<sub>N.inst.plur</sub>

(16) VP[  
     NP.inst[ Нум<sub>N.INST.SING</sub>]  
     V[зачну<sub>V.SING.FUT.PERF</sub>]  
     NP.acc[ NP.acc[наступуюче<sub>A.ACC.SING</sub> поведеніє<sub>N.ACC.SING</sub>]  
             NP.gen[ NP.gen[воєнних<sub>A.GEN.PLUR</sub> діл<sub>N.GEN.PLUR</sub>]  
                     NP.gen[ Хмельницького<sub>N.GEN.SING</sub>]  
                     PP[з Поляки] ] ] ]

Questions to answer:

- What is the processing complexity of this type of structures?
- Can they be accommodated by a finite-state model / context free model?
- Is their processing complexity related to their limited distribution in texts?