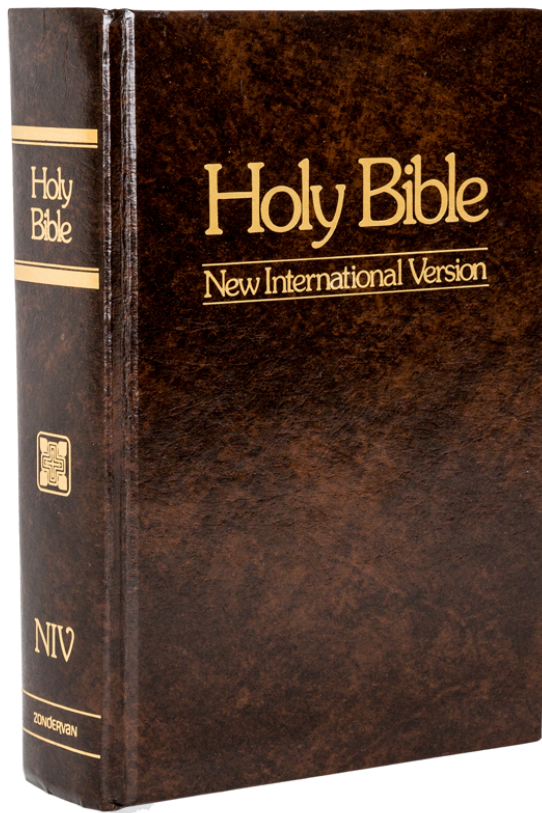


intro



In the beginning was the word. (John1,1)

In the beginning was the word. (John1,1)

Later came the construction.

In the beginning was the word. (John1,1)

Later came the construction.

Dictionaries were made in the old days.

Now we create constructicons instead.

In the beginning was the word. (John1,1)

Later came the **cxn**.

Dictionaries were made in the old days.

Now we create **ccns** instead.

(I will use these abbreviations.)

Show proper respect to everyone. (1Peter2,17)

Show proper respect to everyone. (1Peter2,17)

We love constructions. But we should not forget about words either.

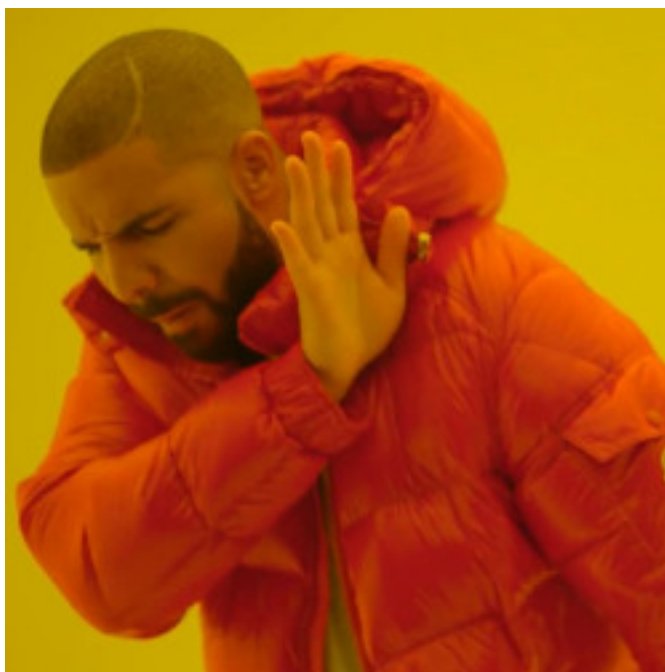
solution: words (even morphemes) are cxns! (cf. Goldberg, 2006, p5)

So ccns will contain everything that dictionaries did.

dictionary $\subset$ **ccn**

my view:

constructicons are not some kind of supplements to dictionaries,
constructicons are **replacements** for dictionaries.



dictionary



constructicon

We need everything: the “ultimate lexical resource” approach to build a constructicon

DICT2CCN workshop
19 April 2024

Bálint Sass



Hungarian Constructicon project – OTKA K 147452

HUN-REN Hungarian Research Centre for Linguistics

Institute for Lexicology

`sass.balint@nytud.hu`

outline

0. intro
1. two approaches
2. the ultimate lexical resource
3. from a dictionary to a ccn
4. demo
5. six aspects
6. availability of ccn-hu
7. summary

approach #1: only „real“ cxns



Janda et al. (2020) writes about the Russian ccn:

“The construction of a language is an open-class inventory that is **potentially limitless. Therefore it would be unrealistic to expect to produce a comprehensive construction resource.** Furthermore, many items [...] are already available in existing reference works, such as **(1) dictionaries** (that contain lexeme-level constructions), **(2) phraseological dictionaries** (that contain idioms where all the slots are fixed), and **(3) grammars** (that explain basic schematic types of sentences and use of function words). **What remains are entrenched multi-word expressions that contain at least one open (not fixed) slot,** and these are the strategic target of the Russian Construction resource.”

→ To create a comprehensive ccn may be impossible,
but it seems worth striving for. **So I choose this path. :)**

approach #2: radically everything



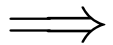
- “all linguistic knowledge consists of units that have both form and meaning, although they vary in terms of schematicity” (CfP)
- **“a person’s knowledge of language consists of nothing but constructions”** (Hilpert, 2014, p2)
- Goldberg (2006, p5) includes bound and free morphemes, not just cxns with open slots

We consider every linguistic form having a meaning a cxn regardless of how complicated or how simple they are.

the ultimate lexical resource

I think that the **right direction of lexicography** is attempting to create an **“ultimate” unified lexical resource** which includes “everything”, i.e. all (kinds of) meaningful linguistic building blocks.

As all meaningful units are cxns,



this ultimate resource is the inventory of cxns, i.e. the ccn. ✓

dictionary $\rightarrow$ ccn

In contrast with other ccn initiatives, which start from languages learner's books and/or corpora, we **start from a monolingual dictionary, and transform it to a ccn.**

As dictionaries attempt to be complete and $\text{dictionary} \subset \text{ccn}$, this decision is consistent with our approach.

transformation:

- lift out MWEs from dictionary entries
- create a separate entry for them
- link them back to the original entry

The *Hungarian Constructicon* project is an attempt to implement this approach and work it out in detail.

demo

similar to Goldberg's Table 1.1

<https://ccn.nytud.hu>

morpheme	' <i>-ban/-ben</i> ' (in)
word	' <i>avokádó</i> ' (avocado)
complex word	' <i>ajándékbolt</i> ' (gift shop)
complex word (semi-filled)	' <i>le-V</i> ' ('down' as a verb particle)
idiom (filled)	' <i>fű·be harap</i> ' ('bite into grass' = bite the dust)
idiom (semi-filled)	' <i>részt vesz X-ban/-ben</i> ' (take part in X)
covariational conditional	' <i>minél X-bb annál Y-bb</i> ' (the X-re the Y-er)
ditransitive	' <i>SUBJ V OBJ1-nak/-nek OBJ2-t</i> '
passive	' <i>SUBJ AUX PP NP_{által}</i> '

six aspects

I will review some of the aspects raised by the current workshop from my “ultimate lexical resource” point of view.

Some are implemented in ccn-hu,
others are future work or just opinions.

1

a) How to define a construction—narrowly or in a broader sense?

- We define the cxn **as broadly as possible**.
- We do not concentrate only on a specific group of linguistic units.
- *include*: morphemes, lexemes,
MWEs with filled slots (phrasemes, sayings, proverbs),
MWEs with open slots, constructional schemas ...

b) Should a construction encompass the most general phrase structure constructions or be limited to idiosyncratic constructions?

- It should.

2

Which constructions should be included in the resource,
i.e., how to arrive at a nomenclature?

- Since we attempt to include “everything”,
→ we do not have to make (often hard) decisions
on whether particular cxns are to be included or not.
- We do not invest into creating/refining a cxn-nomenclature.
We *will* introduce **a very general, dependency-based
cxn-representation** instead, in which all can fit.
- *size*:
~14000 cxns from the MWEs of our initial dictionary
+ about that much expected from processing the examples
ccn-ru has 2200 $\leftrightarrow$ a great amount of fully-fixed cxns in ccn-hu

3

What should a constructional entry look like in the database?

- We are convinced that a cxn entry should be **fully formalized**, fully machine-readable.

How to ensure that the meta-language of descriptions is user-friendly for learners, native speakers, linguists, and NLP applications simultaneously?

- The fully formalized format is suitable for linguists and NLP.
- This should be accompanied with a derived, easily human-readable format for other uses.

4

How to deal with constructions at different levels of abstraction?

- The representation will handle different levels of abstractness.

What relationships and how should be shown between constructions, and between constructions and other units (i.e., lexical) in the same database?

- Cxns will have different kinds of relationships forming a network. E.g. taxonomic, filler-slot, horizontal (see Diessel, 2023, ch3).
- There will be *one* kind of unit in the database, i.e. cxns, so all relationships will be between cxns.
- *important*: our “ultimate lexical resource” approach allows us to include potentially *all* relationships between linguistics units
↔ if we covered only some (“real”) cxns in our resource, inter-resource relationships would be lost!

5

How to fit all constructional information into a data model initially designed for representing lexical units?

- We do not try to fit cxns into a dictionary model!
- We do this the other way round:
the general ccn-representation model, which is to be designed,
will be able to integrate all dictionary lexemes as they are cxns.

6

How can users search for constructions in the dictionary/database? Based on meaning, fixed components, part-of-speech, etc.?

- **analysed search**
→ for querying the “potentially limitless” ccn-hu
- As we want to have a ccn which can tell something about as many linguistic units as possible, we allow **free text as user input and apply morphological (and dependency) analysis** on-the-fly to reveal cxns from the text and present all found cxns to the user.

Analysed search is a very important tool to make the ccn-hu as complete as possible.

availability

<https://ccn.nytud.hu>

username: demo

password: letssee

Feel free to try some queries and
contact me if you have questions or comments.

Bálint Sass

sass.balint@nytud.hu

summary

- There are nothing but constructions.
- Constructicons are replacements for dictionaries.
- *ccn-hu project*: We attempt to make a comprehensive ccn, an ultimate lexical resource, starting from a dictionary.
- Demo: word, suffixed word, cxn.
- Future work: work out a general, dependency-based cxn-representation.
- Analysed search for querying the ccn.
- <https://ccn.nytud.hu> (demo/letssee)

Bálint Sass

sass.balint@nytud.hu

We need everything: the “ultimate lexical resource” approach to build a constructicon

DICT2CCN workshop
19 April 2024

Bálint Sass



Hungarian Constructicon project – OTKA K 147452

HUN-REN Hungarian Research Centre for Linguistics

Institute for Lexicology

`sass.balint@nytud.hu`

References

- Diessel, H. (2023). *The Constructicon*. Cambridge University Press.
- Goldberg, A.E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Hilpert, M. (2014). *Construction Grammar and Its Application to English*. Edinburgh University Press.
- Janda, L., Endresen, A., Zhukova, V., Mordashova, D. & Rakhilina, E. (2020). How to build a constructicon in five years: The Russian example. *Belgian Journal of Linguistics*, 34, pp. 162–175.