So many things to count! Deriving complex numerical expressions in Czech

Marcin Wągiel
Leibniz-Centre General Linguistics (ZAS), Masaryk University in Brno & University of Wrocław

Intro. Most of the research on numerals has focused on the external syntax and semantics of basic cardinals such as *three* in quantifying expressions like *three cats*. However, the recent decade indicated the relevance of investigating the meaning and internal structure of various types of derivationally complex numerical expressions, especially in the context of Slavic data (e.g., Dočekal 2012, Caha 2013, 2017, Wągiel 2015, 2020, 2022, 2023, Wągiel & Dočekal 2018, Khrizman 2020, Dočekal & Grimm 2021; see also Ojeda 1997). Many of such complex expressions give rise to non-trivial semantic effects and show intriguing morphosyntactic idiosyncrasies. Building on previous results, I propose a unified system combining compositional semantics with a late-insertion model grounded in Nanosyntax (Starke 2009) that explains the structure and semantics of complex numerical expressions in Czech. The approach assumes a universal counting mechanism.

Data. Czech distinguishes between several types of numerical expressions, all of which are morphologically complex, as illustrated by the forms in (1)–(6), which all share the root $\sqrt{tr}$- corresponding to the number 3.

(1) tř-i kočky
$\sqrt{3}$-AFF cats
'three cats' BASIC

(2) tr-oj-e klíče
$\sqrt{3}$-AFF$_1$-AFF$_2$ keys
'three sets of keys' AGGREGATE

(3) tr-oj-it-ý hamburger
$\sqrt{3}$-AFF$_1$-AFF$_2$-INFL hamburger
'triple hamburger' MULTIPLIER

(4) tr-oj-í víno
$\sqrt{3}$-AFF$_1$-AFF$_2$ wine
'three kinds of wine' TAXONOMIC

(5) tr-oj-ic-e studentů
$\sqrt{3}$-AFF$_1$-AFF$_2$-INFL students
'group of three students' GROUP

(6) tr-oj-násob-n-ý mistr
$\sqrt{3}$-AFF$_1$-AFF$_2$-ADJ-INFL champion
'three-time champion' FREQUENCY

The make-up and translations of (1)–(6) demonstrate the richness of the Czech system. For the sake of brevity, in the abstract I will illustrate the approach only based on (1)–(3), but it extends to all forms in question. The BASIC numeral (1) consists of the root and the suffix *-i*, which encodes the numeral's ϕ-features. Importantly, it is an idiosyncratic marker that is not found in any other paradigm in Czech. (1) is used to count unspecified entities (by default atomic singular objects). The AGGREGATE numeral (2) consists of the root, morpheme *-oj-* and the suffix *-e*, which again is not a regular inflectional marker. (2) is used not to count singular objects, but rather collections of entities conceptualized as clusters (Grimm & Dočekal 2021). Finally, the MULTIPLIER (3) involves the root, *-oj-*, the suffix *-it-* and a regular marker for adjectival inflection. (3) does not count whole objects, but rather salient parts of a singular entity (Wągiel 2020). Only the basic numeral (1) can be used in the arithmetical function to refer to abstract number concepts, as in 'Three is a prime number' (Wiese 2003, Bultinck 2005, Rothstein 2017). None of the forms in (2)–(6) can express this function. In addition, based on the distribution of the morphemes, the following generalizations in (7) can be formulated.

(7) a. If $\sqrt{tr}$- does not occur with *-oj-*, the following inflectional marker is idiosyncratic (1).
b. If $\sqrt{tr}$- occurs only with *-oj-*, the following inflectional marker is idiosyncratic (2).
c. If $\sqrt{tr}$- occurs with *-oj-* and an additional affix, the following inflectional marker is regular (3).

The morpheme *-oj-* appears also in compounds, in which the numerical component counts entities denoted by the head (8)–(9). Importantly, ϕ-feature marking is absent on the numerical and only appears on the head.

(8) tr-oj-nož-k-a
$\sqrt{3}$-AFF-leg-DIM-INFL
'tripod (lit. three-leg)' NOMINAL COMPOUND

(9) tr-oj-rozměr-n-ý
$\sqrt{3}$-AFF-dimension-ADJ-INFL
'three-dimensional' ADJ. COMPOUND

Semantics. Following Wągiel & Caha (2021), I assume that numerals pronounce complex structures built in a compositional manner from syntactico-semantic primitives. Their underlying meaning is associated with a particular interval on a scale representing the number line. Different types of numerical expressions then encode different kind operations relating to that interval. SCALE (10) denotes a closed interval, i.e., a set of natural numbers, specific for each numeral root (11). NUM (12) is an invariant functional element shared by all numerical expressions, whose meaning is a function from intervals to numbers: MAX yields the greatest value in the interval (13). Finally, a set of related components labeled as different CL heads introduces counting semantics, i.e., they shift a number into a quantifying device that allows for counting some kind of entities via dedicated measure functions (Krifka 1989). # in CL_{OBJ} counts by default singular individuals (14)–(15), ||| in (16) clusters thereof (G&D 2021) and ⊞ in (17) salient parts of a singular object (Wągiel 2020).

(10) $[\![\text{SCALE}_m]\!]_{\langle n,t\rangle} = \lambda n_n[0 \leq n \leq m]$

(11) $[\![\text{SCALE}_3]\!] = [0,3]$

(12) $[\![\text{NUM}]\!]_{\langle\langle n,t\rangle,n\rangle} = \lambda P_{\langle n,t\rangle}[\text{MAX}(P)]$

(13) $[\![\text{NUM}]\!]([\![\text{SCALE}_3]\!]) = 3$

(14) $[\![\text{CL}_{\text{OBJ}}]\!] = \lambda n_n \lambda P_{\langle e,t\rangle}\lambda x_e[{}^*P(x) \wedge \#(P)(x) = n]$

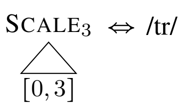
(15) $[\![\text{CL}_{\text{OBJ}}]\!](13) = \lambda P_{\langle e,t\rangle}\lambda x_e[{}^*P(x) \wedge \#(P)(x) = 3]$

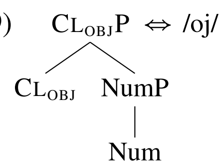
(16) $[\![\text{CL}_{\text{CLSTR}}]\!] = \lambda n_n \lambda P_{\langle e,t\rangle}\lambda x_e[{}^*P(x) \wedge |||(P)(x) = n]$

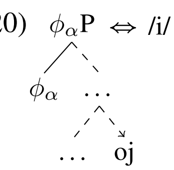
(17) $[\![\text{CL}_{\text{PART}}]\!] = \lambda n_n \lambda P_{\langle e,t\rangle}\lambda x_e[P(x) \wedge \boxplus(P)(x) = n]$

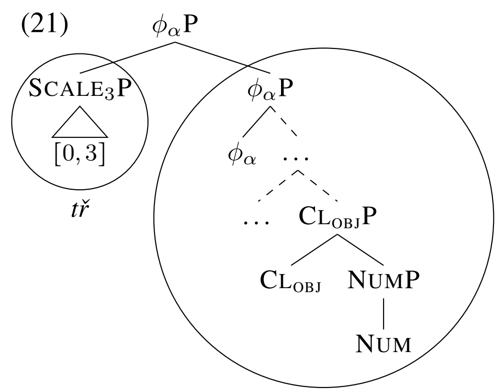
Lexicalization. To capture (1)–(6), I adopt the view that lexical entries link sounds to potentially complex semantic structures. Following a standard nanosyntactic toolbox (Starke 2009 et seq.), I assume the standard cyclic SPELLOUT ALGORITHM and that the SUPERSET PRINCIPLE allows a given morpheme to pronounce *any sub-constituent* contained in its phrasal lexical entry. I also assume that a lexical item can contain a POINTER, which is a node in the structure of an entry that points to another existing entry. Finally, competing candidates for spellout are subject to the ELSEWHERE CONDITION: 'the more specific' entry wins.

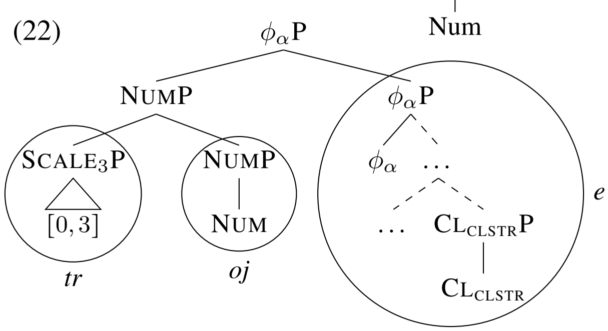
Derivation. First, I propose that Czech numeral roots are stored as (18), i.e., they express only a SCALE component. Next, the morpheme *-oj-* is lexicalized as (19), i.e., it encodes both the number-forging operation NUM and the shift to the default classifier semantics CL_{OBJ}. The suffix *-i* is stored as (20), which involves a pointer to the tree in (19). This means that it can spell out either CL_{OBJ}+NUM or just NUM (due to the Superset Principle) accompanied by the projections pronouncing the relevant ϕ-features. Putting the pieces together in accordance with the Spellout Algorithm results in (21), which expresses the quantifying function of the basic numeral (1), but it is also possible to derive the arithmetical function by spelling out only the NUM part of (19). This is also exactly what happens in (22)–(23), both of which involve different CL components dedicated to counting clusters and parts, respectively. Here, *-oj-* is a better fit and *-e*/*-ý* express ϕ-features.

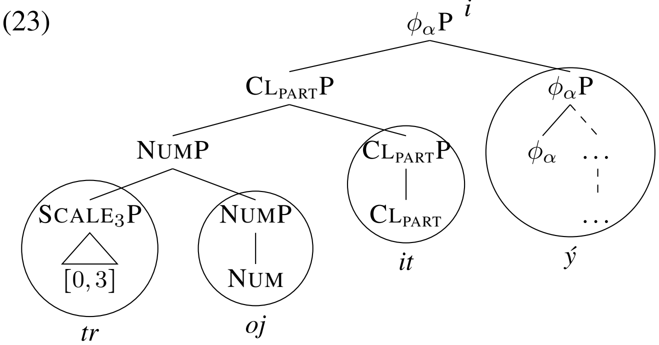
(18)


(19)


(20)


(21)


(22)


(23)


Conclusion. Not only does the proposed account correctly derive the form and meaning of Czech numerical expressions, but also straightforwardly explains the generalizations in (7). The reason why *-i* in (1) and *-e* in (2) are idiosyncratic is because they express not only ϕ-features, but also the respective CL heads. In contrast, *-ý* in (3) is a regular inflectional marker since it only encodes the adjectival ϕ-features and a CL head is pronounced by *-it-*. Finally, the proposal also explains the occurrence of *-oj-* in compounds (8)–(9).

References. Bultinck (2005). *Numerous Meanings* ● Caha (2013) *Czech numerals and no bundling* ● Caha (2017) *Three kinds of 'homogeneous' patterns of Czech numerals* ● Dočekal (2012) *Atoms, groups and kinds in Czech* ● Dočekal & Wągiel (2018) *Event and degree numerals* ● Grimm & Dočekal (2021) *Counting aggregates, groups and kinds* ● Khrizman (2020) *The cardinal/collective alternation in Russian numerals* ● Krifka (1989) *Nominal reference, temporal constitution and quantification in event semantics* ● Ojeda (1997) *A semantics for the counting numerals of Latin* ● Rothstein (2017) *Semantics for Counting and Measuring* ● Starke (2009) *Nanosyntax* ● Wągiel (2015) *Sums, groups, genders, and Polish numerals* ● Wągiel (2020) *Entities, events, and their parts* ● Wągiel (2022) *Quantifying over hidden (parts of) events* ● Wągiel (2023). *Grammatical gender meets classifier semantics* ● Wiese (2003) *Numbers, Language, and the Human Mind*