Underspecification in realisational morphology

Berthold Crysmann and Olivier Bonami

Laboratoire de linguistique formelle — U. Paris Diderot & CNRS

AnaMorphoSys — Lyon, June 2016

Generalisations over exponence

- In many inflection systems, the same exponents may be used in different ways in different contexts.
- We present a formal theory of inflection that is well suited to modeling such situations.
- We highlight 4 types of exponence with variable content:

1. Parallel exponence
   The same shapes realise related but distinct property sets in different positions in the word.
2. Polyfunctionality
   The same shapes realise related but distinct property sets depending on part of speech.
3. Conditioned placement of exponents
   The same shapes realise the same property sets in different positions in different contexts.
4. Gestalt exponence
   Content is assigned to combinations of exponents rather than individual exponents.

Parallel exponence exemplified

- The paradigms of Swahili subject and object markers are nearly identical.

PER	GEN	SUBJECT		OBJECT	
		SG	PL	SG	PL
1		ni	tu	ni	tu
2		u	m	ku	wa
3	M/WA	a	wa	m	wa
	M/MI	u	i	u	i
	KI/VI	ki	vi	ki	vi
	JI/MA	li	ya	li	ya
	N/N	i	zi	i	zi
	U	u	—	u	—
	U/N	u	zi	u	zi
	KU	ku	—	ku	—

Parallel exponence exemplified

- The paradigms of Swahili subject and object markers are nearly identical.
- However, subject and object markers occur in different positions (Stump, 1993).

(1) a. ni-ta-wa-penda
1SG-FUT-3PL-like
'I will like them.'

b. wa-ta-ni-penda
3PL-FUT-1SG-like
'They will like me.'

→ Position, rather than shape, disambiguates which grammatical function is coded.

Polyfunctionality exemplified

- Tundra Nenets uses the same paradigms of person-number and number-case markers in objective conjugation and possessive declension (Ackerman and Bonami, inpress)

(2) a. yemp°q-ŋa-x°yu-da
dress-FIN-DU-3SG
'They two dressed her/him.'

b. ngəno-x°yu-da
boat-DU-3SG
'his/her two boats'

Polyfunctionality exemplified

- Tundra Nenets uses the same paradigms of person-number and number-case markers in objective conjugation and possessive declension (Ackerman and Bonami, inpress)
- This holds even in situations of overlapping exponence

(2) a. meə-m-ʼih
take-SG.1-DU
'We (du.) take it/her/him.'

b. te-m-ʼih
reindeer-NOM.SG.1-DU
'our (du.) reindeer'

- Thus:

Possessed noun~Objective verb
possessor~subject
possessed~object

Conditioned placement exemplified

- In Moro, object markers occur in different positions in different TMA combinations.

(3) a. *g-a-ŋá-vəleð-a*
SM.CL-RTC-2SG.OM-pull-IPFV
's/he is about to pull you' (Jenks and Rose, 2015, 271)

b. *g-á-vəleð-á-ŋá*
SM.CL-DIST.IPFV-pull-DIST.IPFV-2SG.OM
's/he is about to pull you from there to here'

- Object marker placement predictable from tone pattern
- However, a side effect is that the position of object markers acts as secondary exponents of TMA.
- See Crysmann and Bonami (2016) for many more examples and a typology of variable placement.

Gestalt exponence exemplified

- Blevins (2005): while Estonian nouns are easily segmentable, exponents are not associated with stable content.

	‘beak’	
	SG	PL
NOM	nokk	nok-a-d
GEN	nok-a	nokk-a-de
PART	nokk-a	nokk-a-sid

Gestalt exponence exemplified

- Blevins (2005): while Estonian nouns are easily segmentable, exponents are not associated with stable content.
- Stem alternations: {GEN.SG, NOM.PL} vs. all other cells.

	‘beak’	
	SG	PL
NOM	nokk	nok-a-d
GEN	nok-a	nokk-a-de
PART	nokk-a	nokk-a-sid

Gestalt exponence exemplified

- Blevins (2005): while Estonian nouns are easily segmentable, exponents are not associated with stable content.
- Stem alternations: {GEN.SG, NOM.PL} vs. all other cells.
- Theme vowels: NOM.SG vs. all other cells.

	‘beak’	
	SG	PL
NOM	nokk	nok-a-d
GEN	nok-a	nokk-a-de
PART	nokk-a	nokk-a-sid

Gestalt exponence exemplified

- Blevins (2005): while Estonian nouns are easily segmentable, exponents are not associated with stable content.
- Stem alternations: {GEN.SG, NOM.PL} vs. all other cells.
- Theme vowels: NOM.SG vs. all other cells.
- Singular forms contrast in shape, altough no exponent is dedicated to the expression of a particular case value.

	‘beak’	
	SG	PL
NOM	nokk	nok-a-d
GEN	nok-a	nokk-a-de
PART	nokk-a	nokk-a-sid

- “Case properties are realised by the wordforms [...], and words are characterized by different **conbinations** of formatives”.

(Blevins, 2005, 3)

Our goal

- We present aspects of Information-based Morphology, a realisational theory of morphology that embraces the diversity of exponence (Crysmann and Bonami, 2016).
  - In the general case, a realisation rule is a partial generalisation over words linking a set of *m* morphs with a set of *n* morphosyntactic properties.
  - Underspecification allows us to state directly generalisations about exponents at various levels of granularity.
- We show how the theory deals with different types of reuse of exponents.
- We treat two crucial examples:
  1. Parallel exponence in Swahili
  2. Gestalt exponence in Estonian

Important distinctions

1. Constructive vs. abstractive (Blevins, 2006): two modes of description
   - In a constructive approach, the shape of words is deduced from other primitives (morphemes, stems, rules, etc.).
   - In an abstractive approach, words are primitive; stems, exponents, etc. are abstractions deduced from these primitives.
2. Exponence vs. Implicative structure: two empirical questions
   - Exponence is the relation between properties expressed by a word and aspects of the word's shape expressing them.
   - Implicative relations are relations between words expressing different property sets.

Important distinctions

- Classical generative morphology is a constructive approach to exponence.
- Blevins (2006); Ackerman et al. (2009) and the following literature adopt an abstractive approach to implicative relations.
- We argue that the two distinctions are orthogonal.
- The present approach:
  - has both constructive and abstractive interpretations;
  - is entirely focused on exponence.

Realisations rules as generalisations over words I

- For the purposes of inflection, words can be seen as associations between a phonological shape (PH) and a morphosyntactic property set (MS).

$$\begin{bmatrix} \text{PH} & \text{<ɹeɪnɪŋ>} \\ \text{MS} & \left\{ \begin{bmatrix} \text{LID} & \textit{rain} \end{bmatrix}, \begin{bmatrix} \text{TMA} & \textit{prs-ptcp} \end{bmatrix} \right\} \end{bmatrix}$$

- As a first approximation, rules of exponence can be seen as underspecified descriptions of words.

$$\begin{bmatrix} \text{PH} & \text{<...ɪŋ>} \\ \text{MS} & \left\{ \begin{bmatrix} \text{TMA} & \textit{prs-ptcp} \end{bmatrix}, \ldots \right\} \end{bmatrix}$$

Realisations rules as generalisations over words II

- Because words can consist of more than two bits, we need some way to index position within a word.
  → rule blocks in AMM (Anderson, 1992) and PFM (Stump, 2001)
- Instead we use explicit reference to numbered positions.
  → explicit list of morphs (MPH)

Word:

$$\begin{bmatrix} \text{PH} & \text{<ɹeɪnɪŋ>} \\ \text{MPH} & \left\{ \begin{bmatrix} \text{PH} & \text{<ɹeɪn>} \\ \text{PC} & 0 \end{bmatrix}, \begin{bmatrix} \text{PH} & \text{<ɪŋ>} \\ \text{PC} & 1 \end{bmatrix} \right\} \\ \text{MS} & \left\{ \begin{bmatrix} \text{LID} & \textit{rain} \end{bmatrix}, \begin{bmatrix} \text{TMA} & \textit{prs-ptcp} \end{bmatrix} \right\} \end{bmatrix}$$

Rule of exponence:

$$\begin{bmatrix} \text{MPH} & \left\{ \begin{bmatrix} \text{PH} & \text{<ɪŋ>} \\ \text{PC} & 1 \end{bmatrix}, \ldots \right\} \\ \text{MS} & \left\{ \begin{bmatrix} \text{TMA} & \textit{prs-ptcp} \end{bmatrix}, \ldots \right\} \end{bmatrix}$$

- Trivial relationship between a word's phonology (a string) and its morphs (a set of strings indexed for position).
- Easily captures cumulative exponence (1 morph:*n* properties), extended exponence (*m*:1) and overlapping exponence (*m*:*n*).

Realisations rules as generalisations over words III

- However, this simple view does not allow one to speak of situations where the same association between form and content is used more than once in the same word.
  - Parallel exponence (see above)
  - Exuberant exponence (Harris, 2009)
- We add an extra layer of abstraction:
  1. A word's representation includes a specification of which realisation rules license the relation between its form and content.
  2. Realisation rules express a relation between a set of morphs of fixed arity and a specific set of morphosyntactic properties, the morphology under discussion (MUD).

$$\begin{bmatrix} \text{MPH} & \left\{ \begin{bmatrix} \text{PH} & \langle \imath\eta \rangle \\ \text{PC} & 1 \end{bmatrix} \right\} \\ \text{MUD} & \left\{ \begin{bmatrix} \text{TMA} & \textit{prs-ptcp} \end{bmatrix} \right\} \end{bmatrix}$$

  3. A principle of morphological well-formedness ensures that
     - 3.1 The properties expressed by rules add up to the word's property set
     - 3.2 The morphs introduced by rules add up to the word's morph list.

Realisations rules as generalisations over words IV

- For the technically inclined:

$$word \rightarrow \begin{bmatrix} \textsc{mph} & \boxed{e_1} \cup \cdots \cup \boxed{e_n} \\ \textsc{rr} & \left\{ \begin{bmatrix} \textsc{mph} & \boxed{e_1} \\ \textsc{mud} & \boxed{m_1} \end{bmatrix}, \ldots, \begin{bmatrix} \textsc{mph} & \boxed{e_n} \\ \textsc{mud} & \boxed{m_n} \end{bmatrix} \right\} \\ \textsc{ms} & \boxed{m_1} \uplus \cdots \uplus \boxed{m_n} \end{bmatrix}$$

Morphological well-formedness

- In our example:

$$\begin{bmatrix} \textsc{mph} & \left\{ \begin{bmatrix} \textsc{ph} & \text{<ɹeɪn>} \\ \textsc{pc} & 0 \end{bmatrix}, \begin{bmatrix} \textsc{ph} & \text{<ɪŋ>} \\ \textsc{pc} & 1 \end{bmatrix} \right\} \\ \textsc{rr} & \left\{ \begin{bmatrix} \textsc{mph} & \left\{ \begin{bmatrix} \textsc{ph} & \text{<ɹeɪn>} \\ \textsc{pc} & 0 \end{bmatrix} \right\} \\ \textsc{mud} & \left\{ \begin{bmatrix} \textsc{lid} & \textit{rain} \end{bmatrix} \right\} \end{bmatrix}, \begin{bmatrix} \textsc{mph} & \left\{ \begin{bmatrix} \textsc{ph} & \text{<ɪŋ>} \\ \textsc{pc} & 1 \end{bmatrix} \right\} \\ \textsc{mud} & \left\{ \begin{bmatrix} \textsc{tma} & \textit{prs-ptcp} \end{bmatrix} \right\} \end{bmatrix} \right\} \\ \textsc{ms} & \left\{ \begin{bmatrix} \textsc{lid} & \textit{rain} \end{bmatrix}, \begin{bmatrix} \textsc{tma} & \textit{prs-ptcp} \end{bmatrix} \right\} \end{bmatrix}$$

Realisations rules as generalisations over words V

- In short:
  - Realisation rules are abstractions over words, stating that some collection of morphs jointly express some collection of properties.
  - Morphological well-formedness ensures ‘Total Accountability’ (Hockett, 1947).
  - The 1:1 relation of the classical morpheme is one possibility, but the framework accomodates many other situations.

Generalisations over rules

- Back to our initial goal: capturing the variable content of exponents.
- Example: Swahili

(4) a. ni-ta-wa-penda
1SG-FUT-3PL-like
'I will like them.'

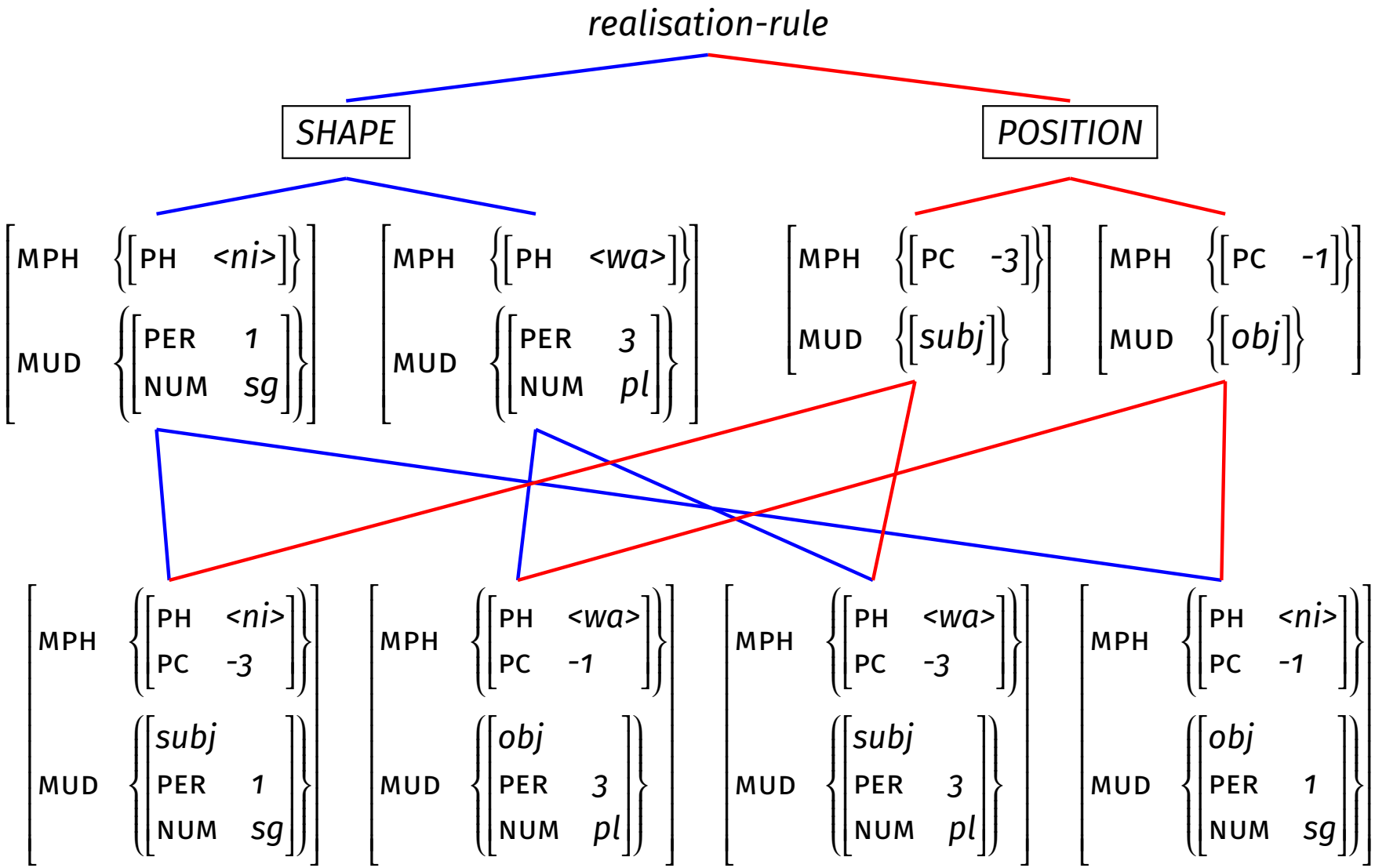
$$\left[\begin{array}{ll}\textsc{mph} & \left\{\left[\begin{array}{ll}\textsc{ph} & \langle ni\rangle\\ \textsc{pc} & -3\end{array}\right]\right\}\\ \textsc{mud} & \left\{\left[\begin{array}{ll}subj & \\ \textsc{per} & 1\\ \textsc{num} & sg\end{array}\right]\right\}\end{array}\right]\left[\begin{array}{ll}\textsc{mph} & \left\{\left[\begin{array}{ll}\textsc{ph} & \langle wa\rangle\\ \textsc{pc} & -1\end{array}\right]\right\}\\ \textsc{mud} & \left\{\left[\begin{array}{ll}obj & \\ \textsc{per} & 3\\ \textsc{num} & pl\end{array}\right]\right\}\end{array}\right]$$

b. wa-ta-ni-penda
3PL-FUT-1SG-like
'They will like me.'

$$\left[\begin{array}{ll}\textsc{mph} & \left\{\left[\begin{array}{ll}\textsc{ph} & \langle wa\rangle\\ \textsc{pc} & -3\end{array}\right]\right\}\\ \textsc{mud} & \left\{\left[\begin{array}{ll}subj & \\ \textsc{per} & 3\\ \textsc{num} & pl\end{array}\right]\right\}\end{array}\right]\left[\begin{array}{ll}\textsc{mph} & \left\{\left[\begin{array}{ll}\textsc{ph} & \langle ni\rangle\\ \textsc{pc} & -1\end{array}\right]\right\}\\ \textsc{mud} & \left\{\left[\begin{array}{ll}obj & \\ \textsc{per} & 1\\ \textsc{num} & sg\end{array}\right]\right\}\end{array}\right]$$

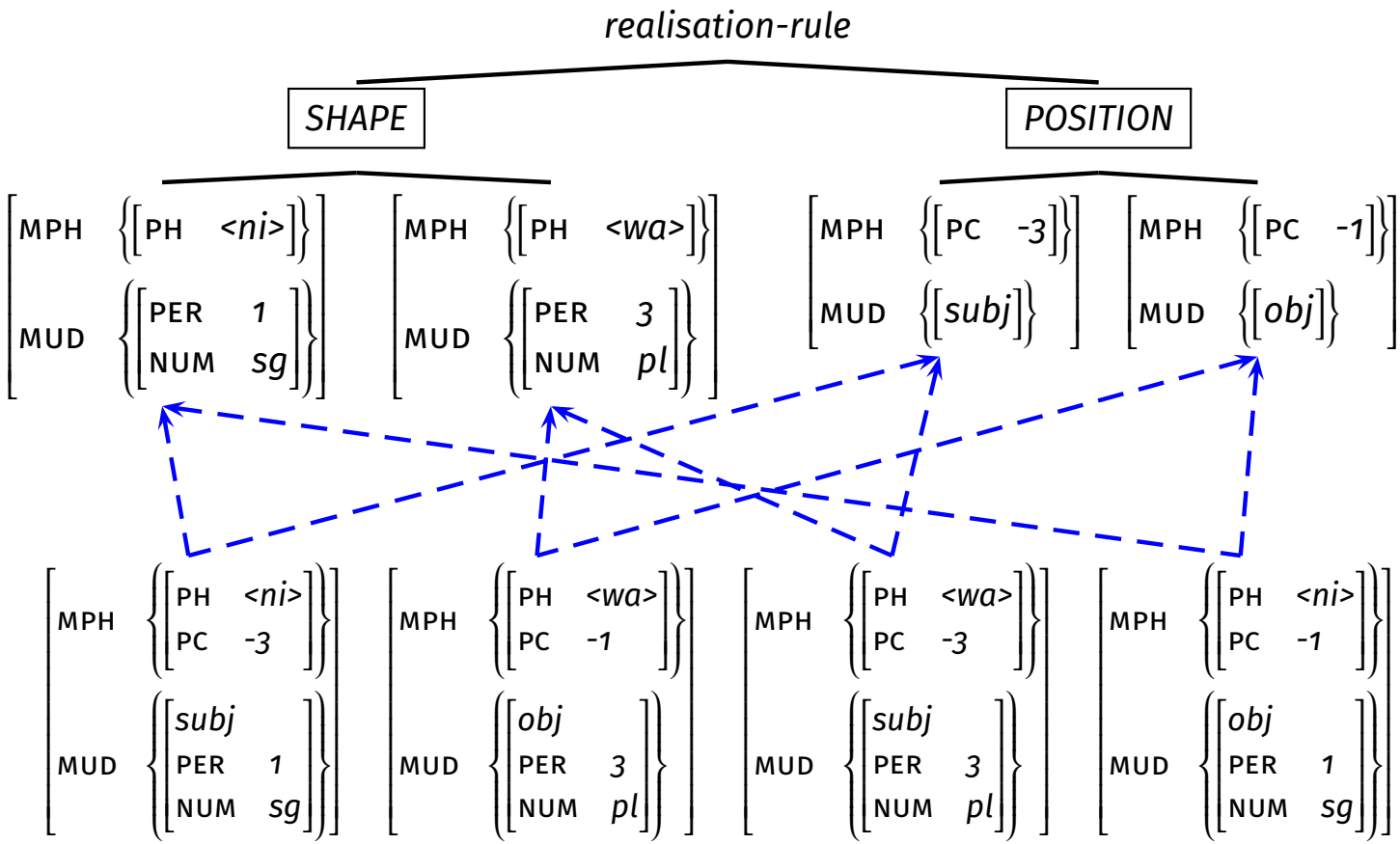
Hierarchies of rules

- Strategy familiar from HPSG: organise realisation rules into a (monotonous) multiple inheritance hierarchy



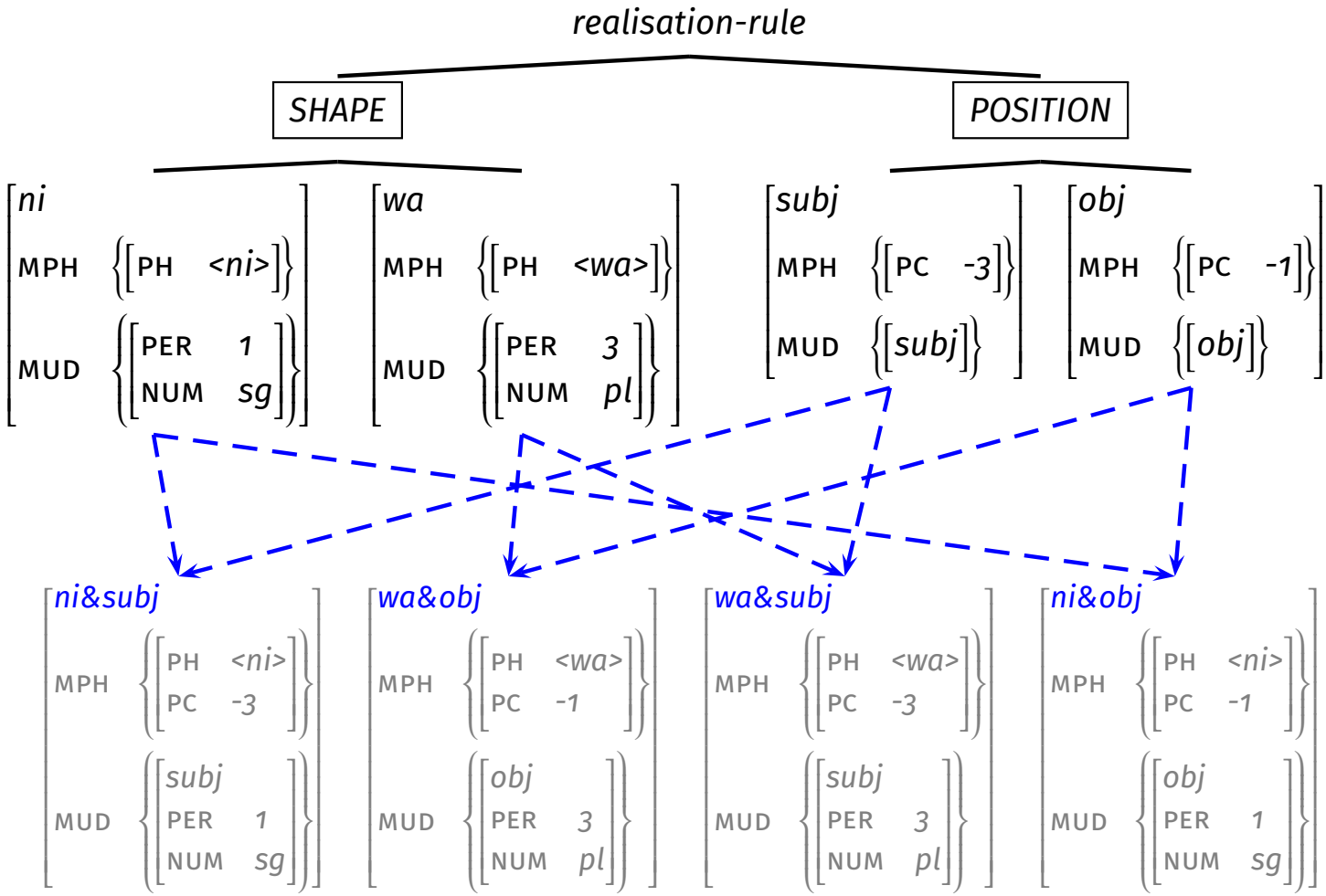
Hierarchies of rules

- Monotonous multiple inheritance hierarchies have a natural abstractive interpretation: nodes in the hierarchy state what some words (or word parts) have in common.



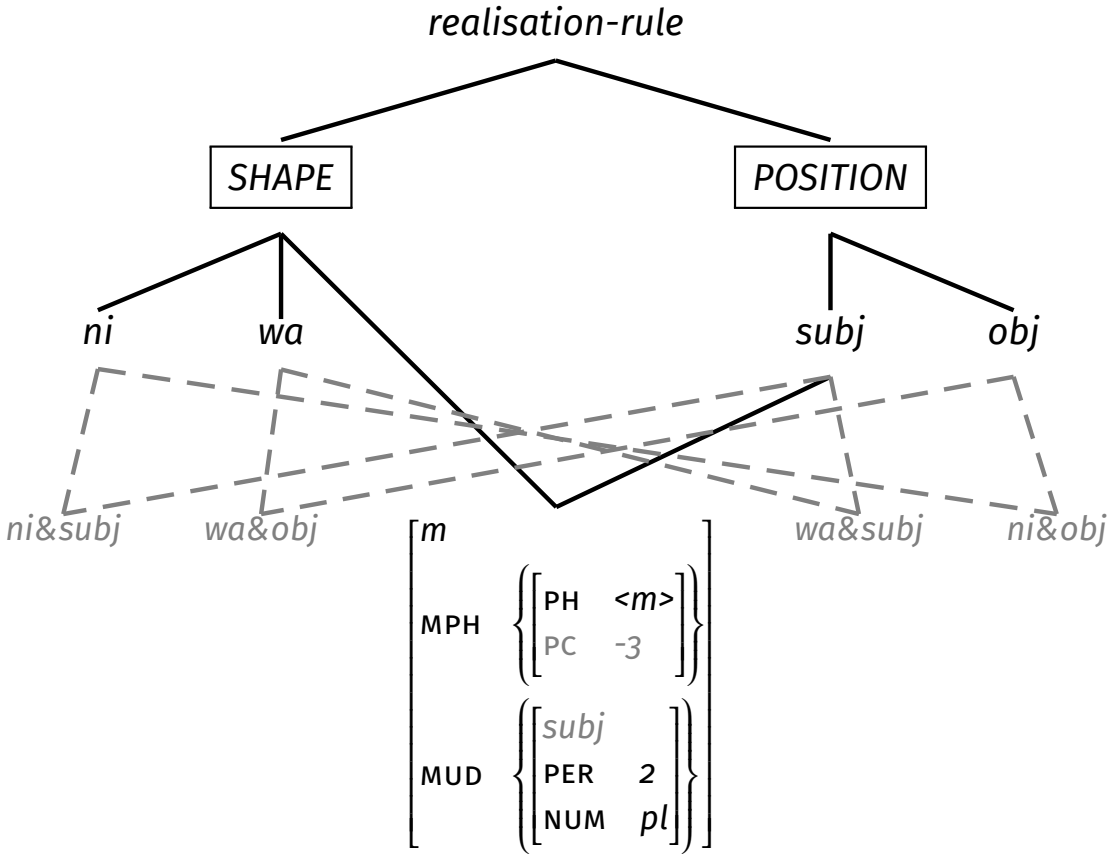
Hierarchies of rules

- A constructive interpretation of the same hierarchies can be given using online type construction (Koenig and Jurafsky, 1994).
- The complete hierarchy is deduced from a reduced hierarchy by expanding all combinations of types.



Hierarchies of rules

- Pre-linking a rule in multiple dimensions blocks overgeneralisation.



Interim conclusion

- We present a view of exponence where:
  - A single rule may link *m* properties with *n* exponents
  - Similarities and differences between rules are captured in a monotonous multiple inheritance hierarchy
  - Because it is monotonous and multi-dimensional, the hierarchy can be interpreted abstractively or constructively.
- Allows for a simple account of parallel exponence in Swahili.
- For Swahili, it is crucial that exponents of subject and object marking be introduced separately
  - This allows us to say that rules for subjects and objects have something in common
- We now turn to a system where it is crucial that all exponents be introduced simultaneously.

Back to Estonian

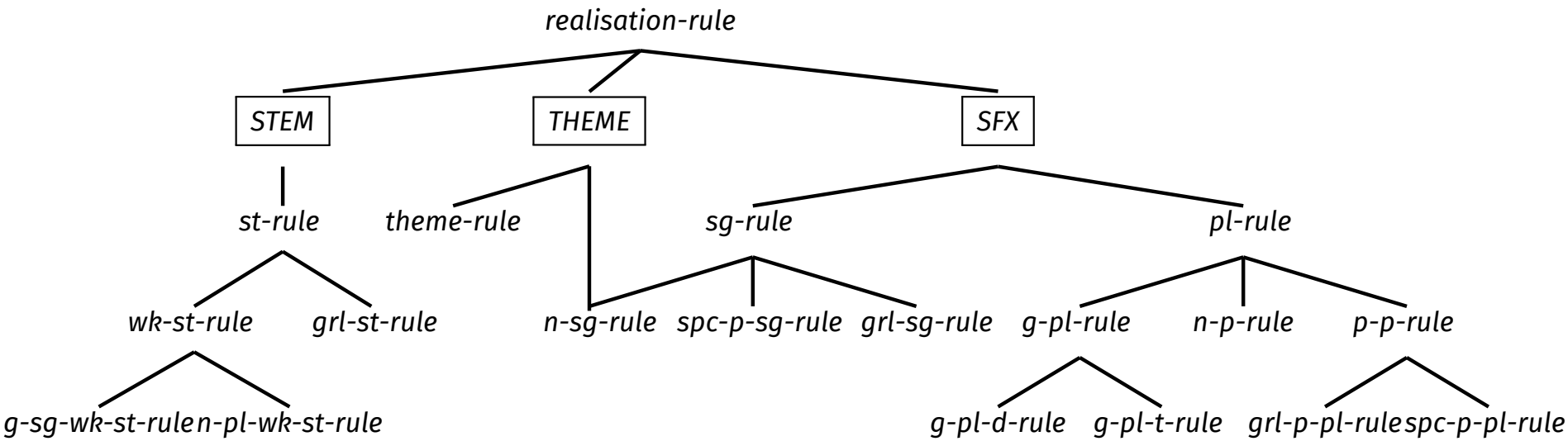
- In Estonian declension, the number of morphs in a word plays a crucial role in exponence.

	'beak'		'workbook'		'seminar'	
	SG	PL	SG	PL	SG	PL
NOM	nokk	nok-a-d	õpik	õpik-u-d	seminar	seminar-i-d
GEN	nok-a	nokk-a-de	õpik-u	õpik-u-te	seminar-i	seminar-i-de
PART	nokk-a	nokk-a-sid	õpik-u-t	õpik-u-id	seminar-i	seminar-i-sid

- In these inflection classes:
  - The plural is characterised by the presence of 3 distinct morphs
  - 1 to 3 morphs in the singular.
  - The nominative singular is characterised by a bare stem
- This motivates a holistic analysis, where all morphs in a word jointly realize content.
- Can be readily captured in the present framework.

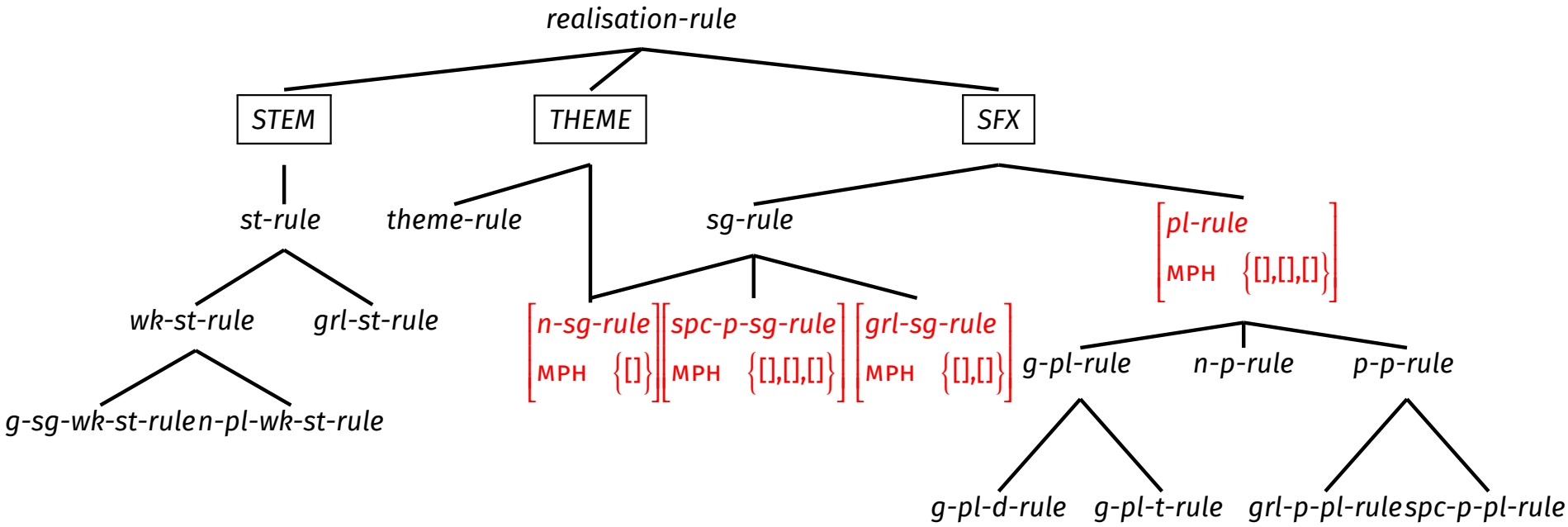
Simultaneous introduction in Estonian

- Three dimensions controlling:
  - STEM the choice of a stem alternant
  - THEME the possible introduction of a theme vowel
  - SFX the possible introduction of a case-number suffix

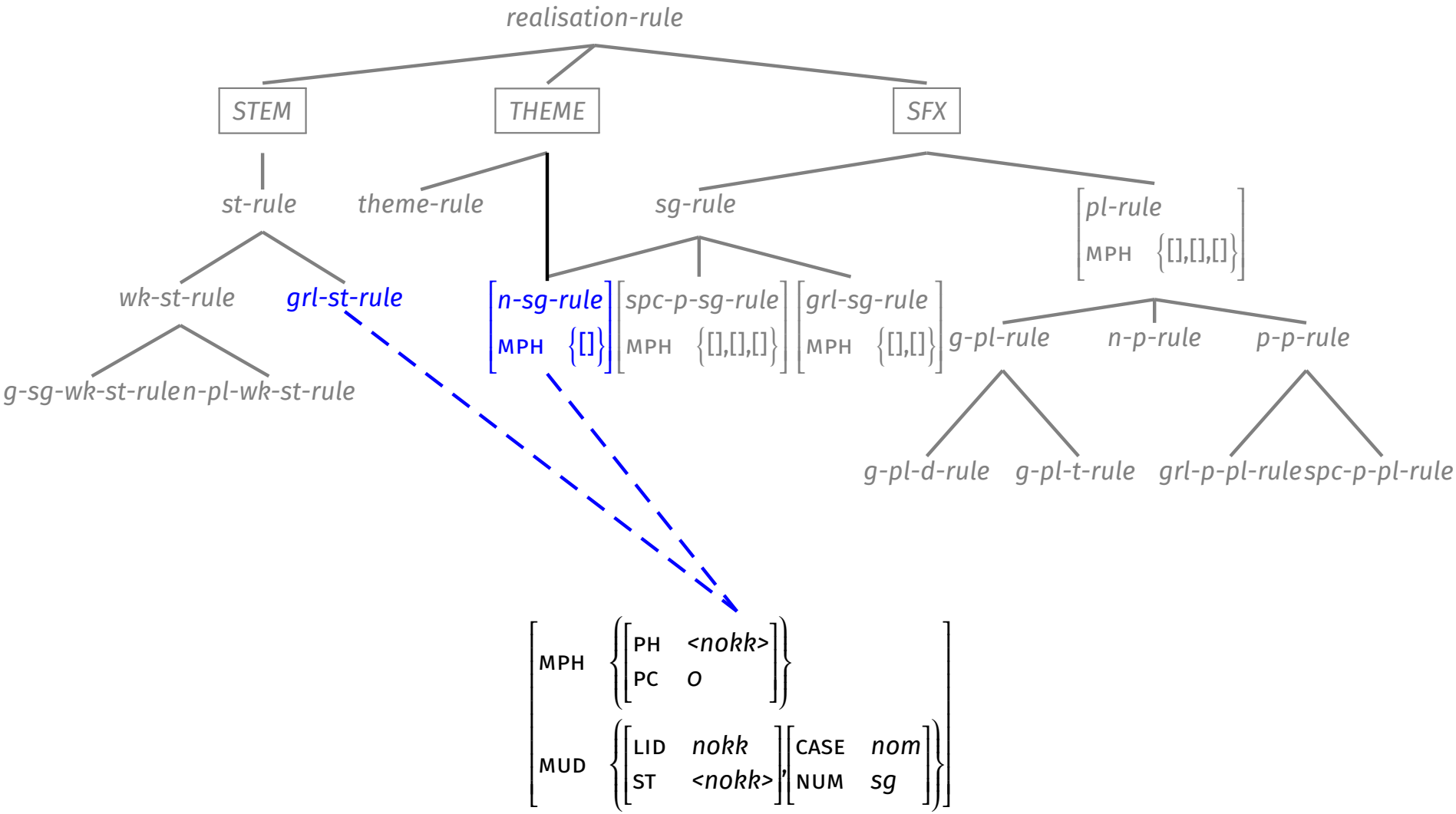


Simultaneous introduction in Estonian

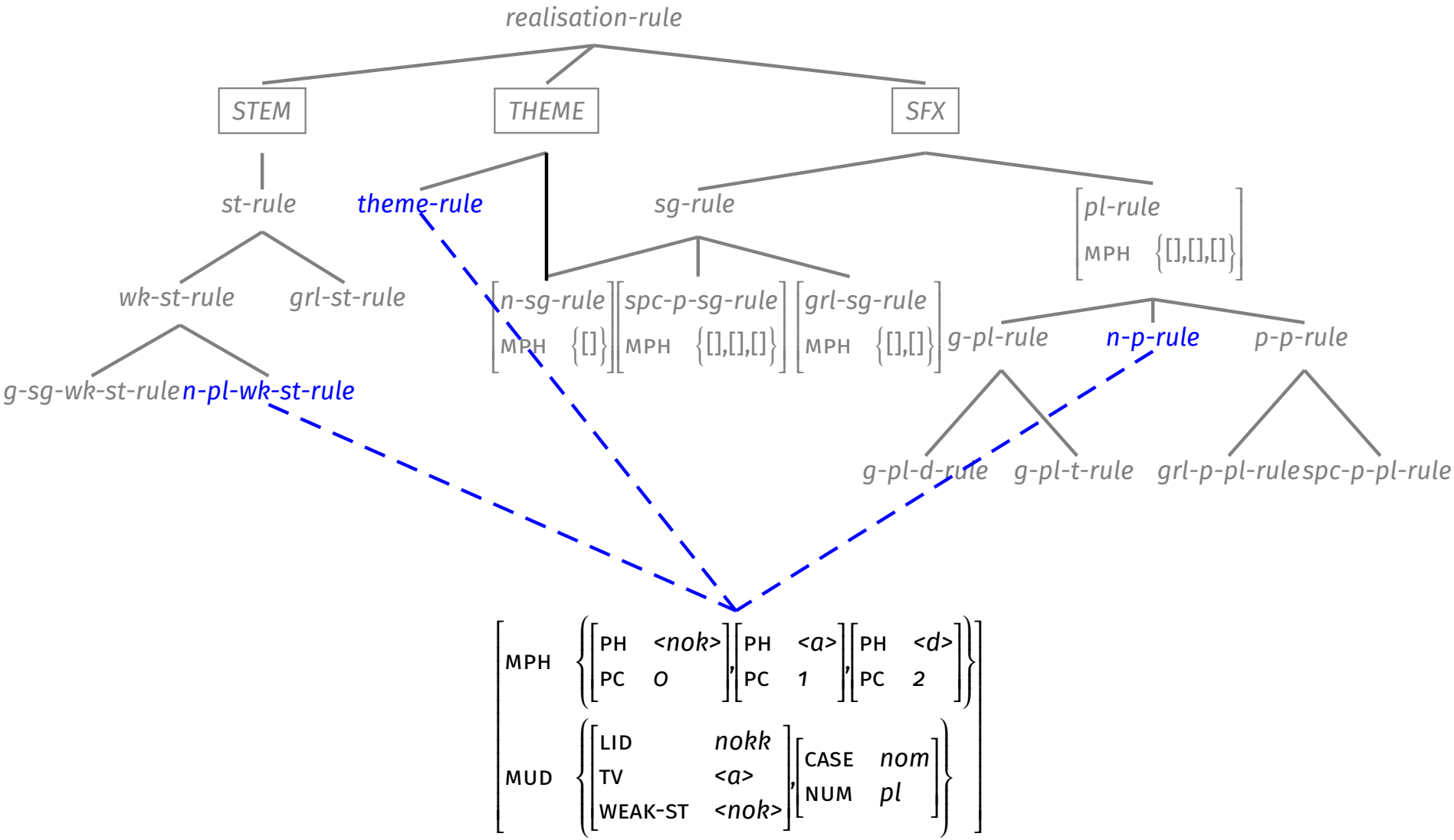
- Some rule types in the THEME and SFX dimensions jointly determine the arity of the set of morphs:



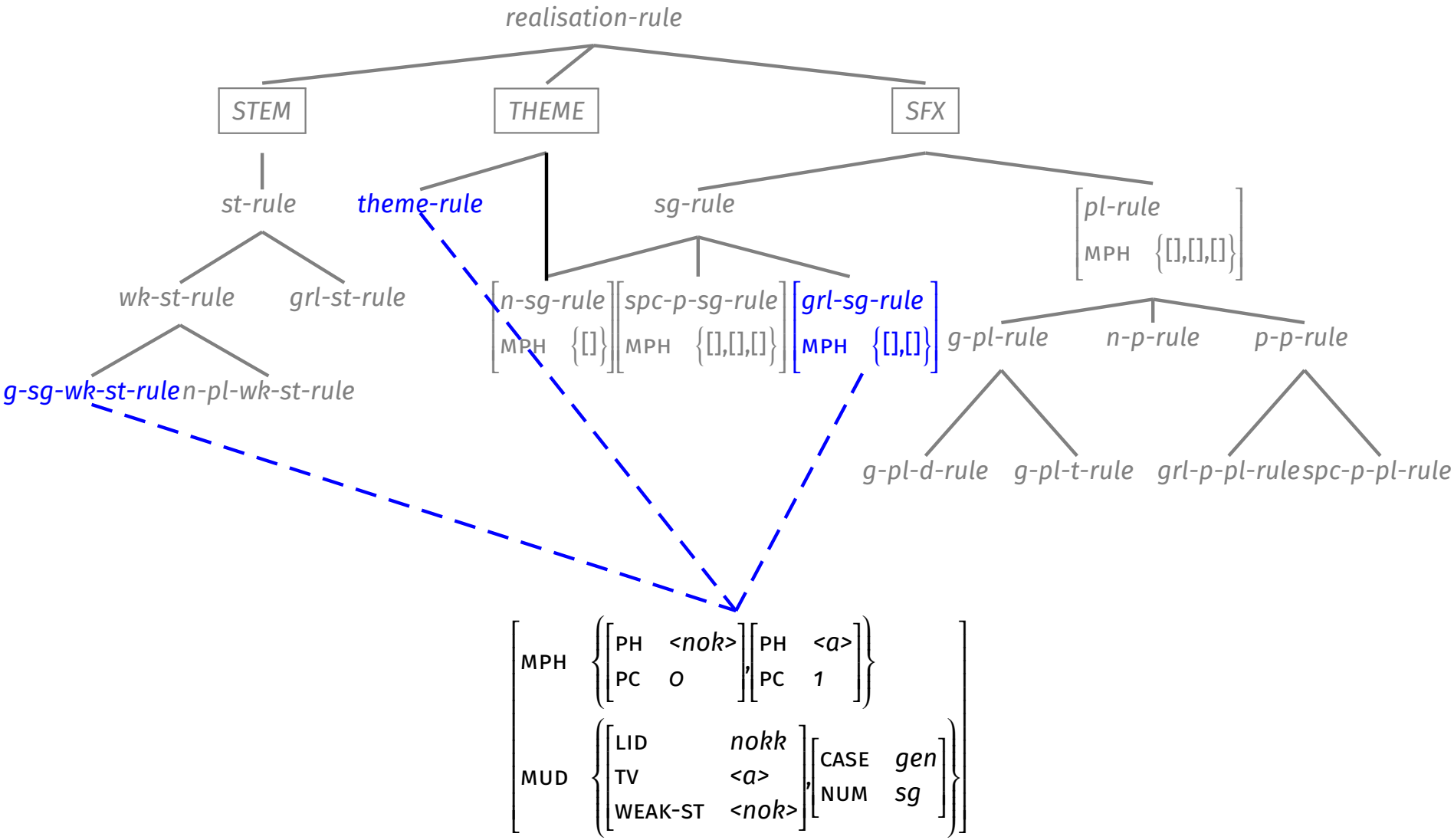
Simultaneous introduction in Estonian



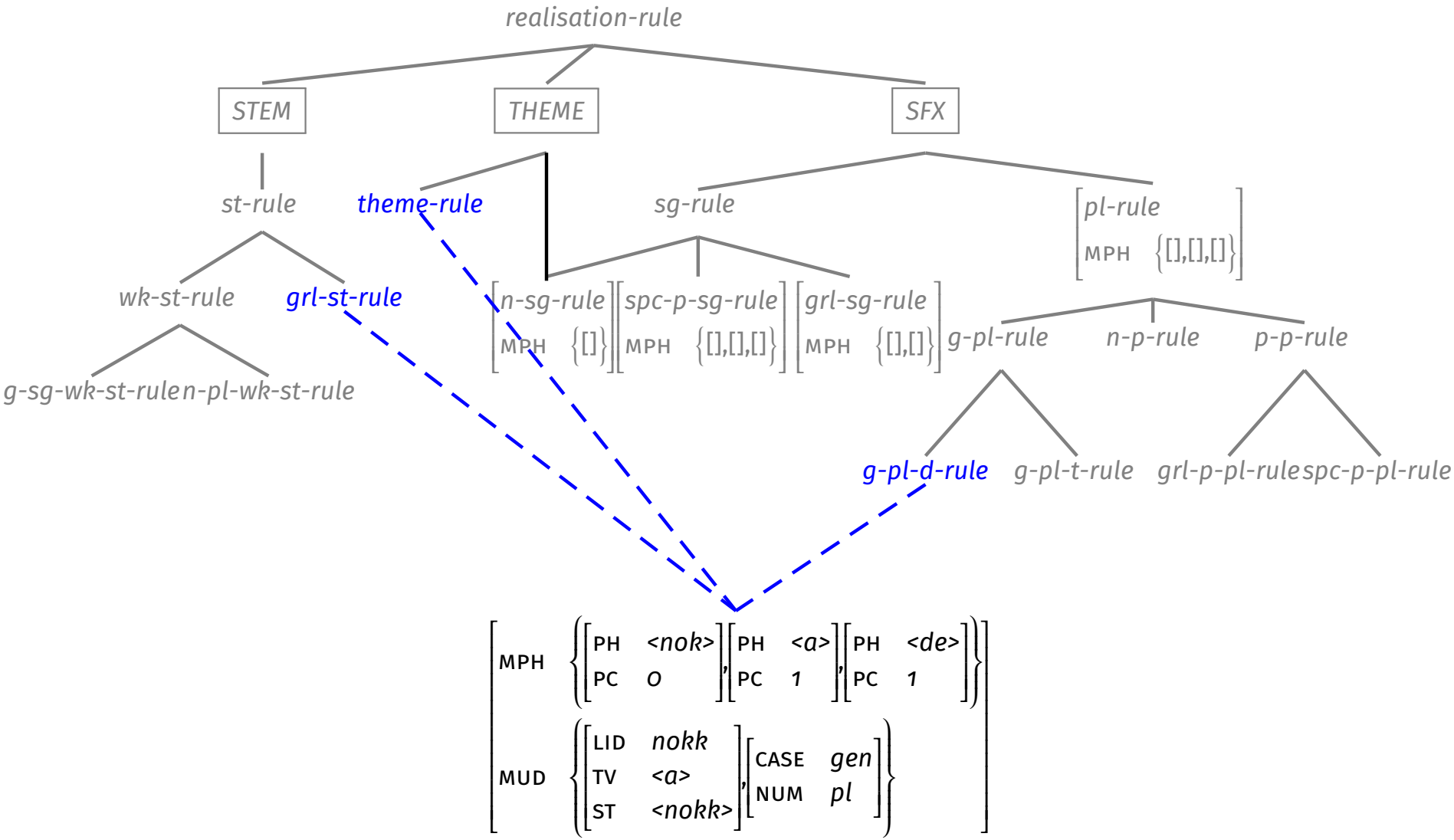
Simultaneous introduction in Estonian



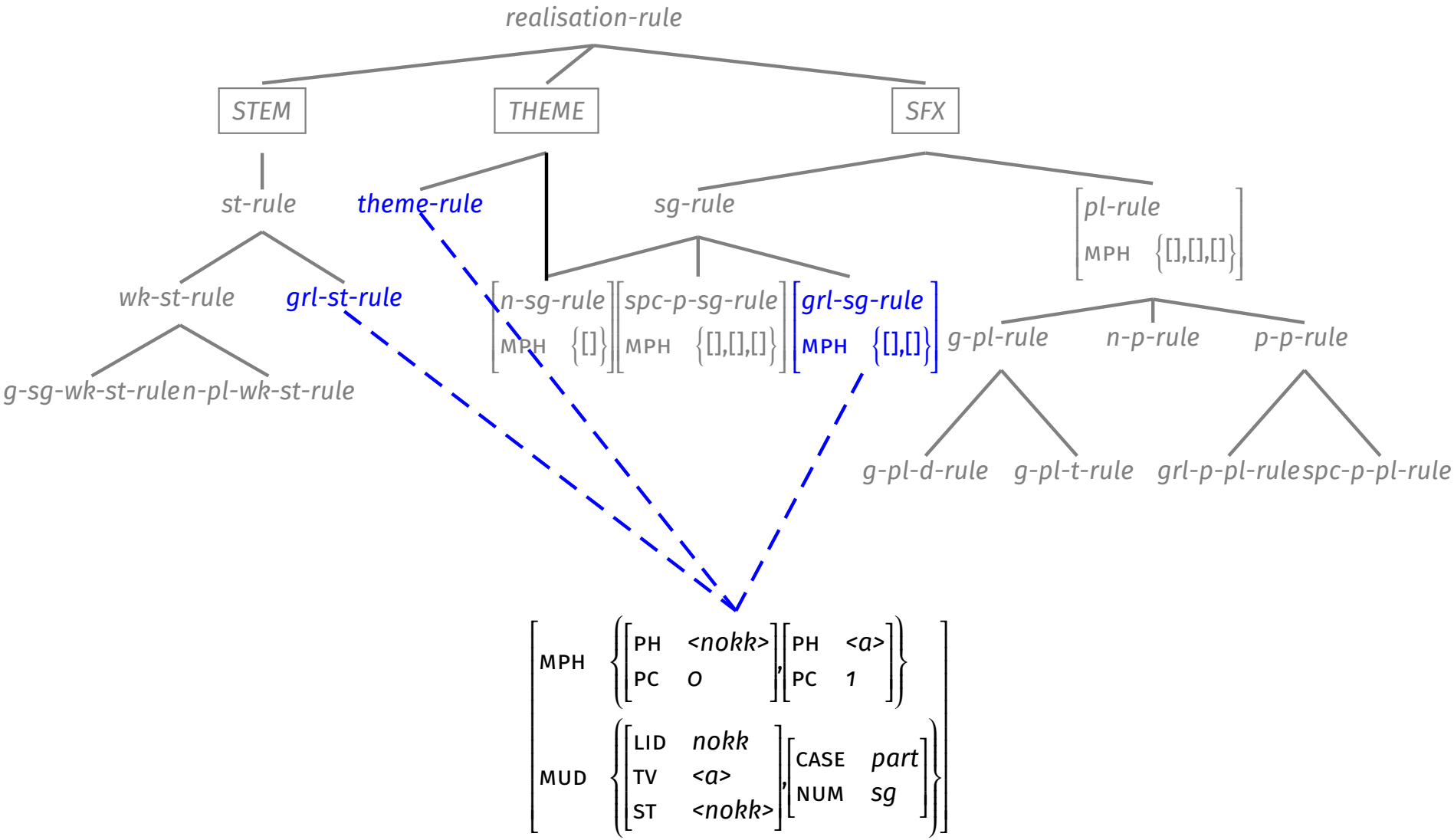
Simultaneous introduction in Estonian



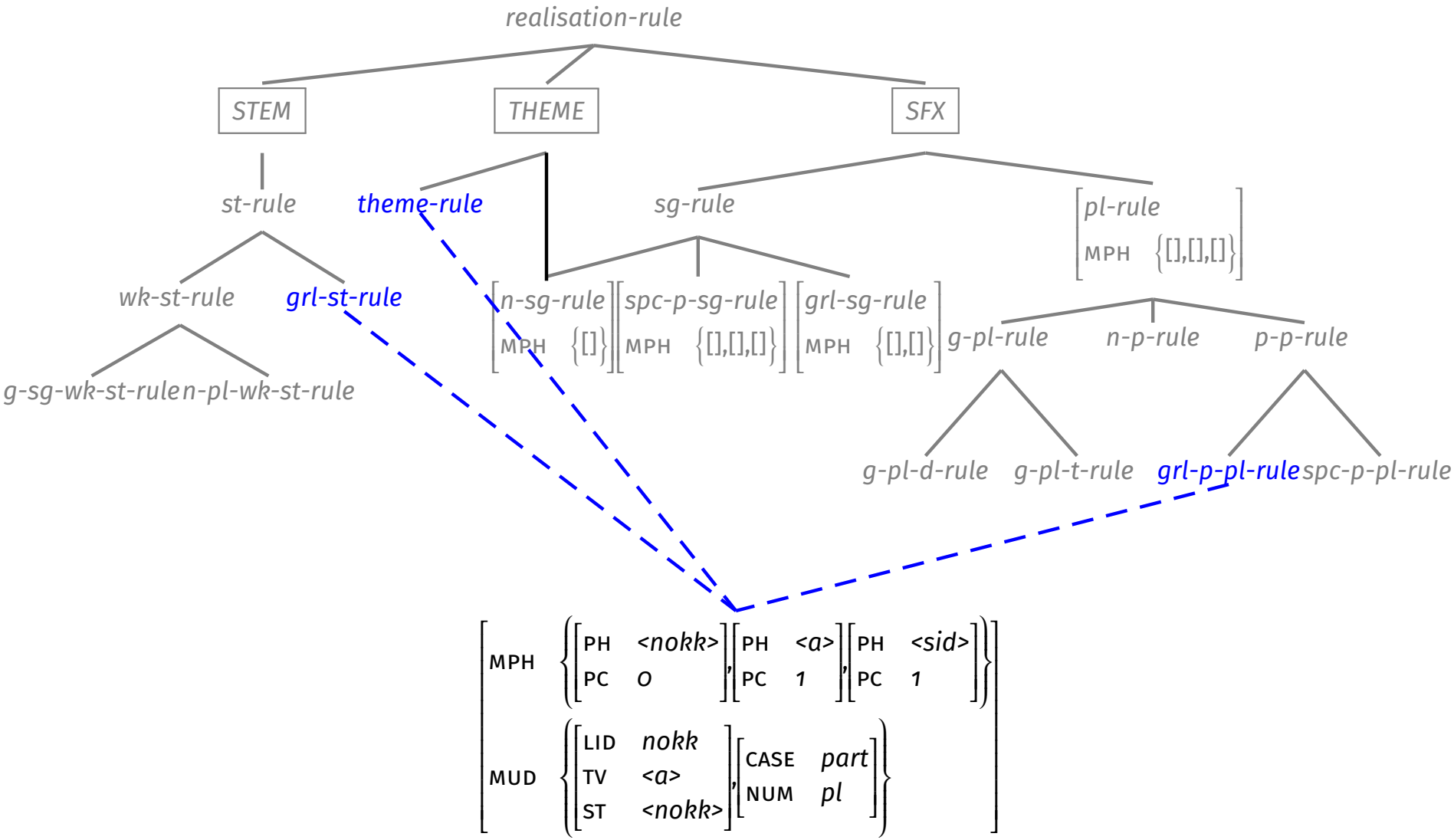
Simultaneous introduction in Estonian



Simultaneous introduction in Estonian



Simultaneous introduction in Estonian



Conclusions on Estonian

- This account captures crucial insights of Blevins (2005); Blevins et al. (in press) on the Estonian declension system:
  - Segmentation is clear, but there is no stable association between segments and morphosyntactic content
  - Each dimension captures a series of contrasts, although these contrasts are not stictly tied to positions.
  - Paradigmatic opposition is captured holistically for the word
    - No empty element is needed.
- But:
  - The account can be made sense of both in constructive and in abstractive terms.
  - The account says nothing on implicative relations
    - This is deliberate: we take exponence and implicative structure to be orthogonal questions.

Conclusions

- Exponents with variable content should be a core concern of theories of inflection.
- Information-based Morphology is particularly well-equipped to address such situations:
  - Individual rules express *m*:*n* relations between form and content.
  - Underspecification as a single mechanism to capture similarity.
- Two case studies:
  - A proper treatment of Swahili requires individual introduction of exponents
  - A proper treatment of Estonian requires holistic introduction of exponents.
- We provide a formally sound basis for developing a constructional approach to inflection (Gurevich, 2006).
  - Rules of exponence are word-internal constructions
    - organized in a system of paradigmatic oppositions,
    - ranging from the most specific to the most abstract.
  - The combinatorics are very different from that of *syntactic* constructions.

References

Ackerman, F., Blevins, J. P., and Malouf, R. (2009). 'Parts and wholes: implicative patterns in inflectional paradigms'. In J. P. Blevins and J. Blevins (eds.), *Analogy in Grammar*. Oxford: Oxford University Press, 54–82.

Ackerman, F. and Bonami, O. (inpress). 'Systemic polyfunctionality and morphology-syntax interdependencies'. In A. Hippisley and N. Gisborne (eds.), *Defaults in Morphological Theory*. Oxford: Oxford University Press.

Anderson, S. R. (1992). *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Blevins, J. P. (2005). 'Word-based declensions in Estonian'. In G. E. Booij and J. v. Marle (eds.), *Yearbook of Morphology 2005*. Springer, 1–25.

——— (2006). 'Word-based morphology'. *Journal of Linguistics*, 42:531–573.

Blevins, J. P., Ackerman, F., and Malouf, R. (in press). 'Morphology as an adaptive discriminative system'. In H. Harley and D. Siddiqi (eds.), *Morphological Metatheory*. Amsterdam: John Benjamins, 35pp.

Crysmann, B. and Bonami, O. (2016). 'Variable morphotactics in Information-Based Morphology'. *Journal of Linguistics*, 52:311–374.

Gurevich, O. I. (2006). *Construction Morphology: the Georgian version*. Ph.D. thesis, University of California, Berkeley.

Harris, A. C. (2009). 'Exuberant exponence in Batsbi'. *Natural Language and Linguistic Theory*, 27:267–303.

Hockett, C. F. (1947). 'Problems of morphemic analysis'. *Language*, 23:321–343.

Jenks, P. and Rose, S. (2015). 'Mobile object markers in Moro: The role of tone'. *Language*, 91:269–307.

Koenig, J.-P. and Jurafsky, D. (1994). 'Type underspecification and on-line type construction'. In *Proceedings of WCCFL XIII*. 270–285.

Rose, S. (2013). 'The morphological structure of the Moro verb'. In R. Blench and T. Schadeberg (eds.), *Nuba Mountain Language Studies*. Cologne: Rüdiger Köppe, 25–56.

Stump, G. T. (1993). 'Position classes and morphological theory'. In G. E. Booij and J. van Marle (eds.), *Yearbook of Morphology 1992*. Kluwer, 129–180.

——— (2001). *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.