Underspecification of nominal functional categories in Semitic and Slavic¹

1 Cross-linguistically parallel homophony?

In Slavic:

- the default diminutive (DIM) morpheme (henceforth, \mathbf{K} ; e.g., Czech -ek.M.SG, -ka.F.SG, -ko.N.SG etc.)² can yield a degree interpretation, and obtain additional pragmatic readings
- but it is also *homophonous* with other functional morphemes (a nominalizer, a conceptual³ female-denoting, and a group-forming morpheme)

A very similar range of nominal functions and interpretations is found in *Semitic*:

- ⇒ the feminine morpheme (**F**) displays a similar range of functional and semantic interpretations, e.g., in Moroccan and Levantine Arabic (LA), and Hebrew, with two important differences:
 - F individuates, and
 - F does not realize a nominalizing head

How does functional/interpretational variability within the nominal domain map to PF uniformity?

- homophony over a number of functional interpretations within a single language, or even a single family of languages, is not surprising in and of itself
- but *parallel systematic* homophony over the same set of functional interpretations, and structural restrictions on their syntactic behaviour and distributional/functional gaps *across language families* requires a structural explanation

We argue against an accidental homophony:

• instead, both **K** and **F** are morphological realizations of a feature bundle corresponding to a *underspecified nominal functional head* (*i**),⁴ construed as a variable (in the sense of Borer 2005 and Acquaviva 2018)

¹We are indebted to Aya Zarka for the Levantine Arabic data and for her insights and discussions about the language. We thank to the Social Sciences and Humanities Research Council of Canada for funding this research (SSHRC Insight Grant #435-2016-1034, Grammatical vs semantic features: the semantics-morphology mapping, and its consequences for syntax; PI: Kučerová).

²Slavic languages display a range of morphologically distinct and lexically specified derivational morphemes expressing a diminutive-like meaning (see, for example, Steriopolo 2009, 2013; Wiltschko & Steriopolo 2008; Krizhman 2019). Here, we are concerned only with the default and fully productive form as this is the only morpheme that exhibits the functional variability in the centre of our research investigation.

³We use the term 'conceptual gender' to refer to what the older literature calls 'natural,' 'biological' or 'sexbased' gender. See, e.g., Ackerman 2019 for an argument why the terminological change better reflects the intended denotation.

 $^{^4}$ The i^* notation is based on the i^* heads Wood & Marantz (2015) according to whom i^* 's functional interpretation is assigned at the interface based on their syntactic configuration. In the present proposal, the functional interpretation is established within narrow syntax.

- where the range of their functional properties is a function of their syntactic position
- \Rightarrow different functions and interpretations arise from different attachments sites of i^* in the extended nominal domain, instead of a series of semantically specified functional heads (e.g., Fassi Fehri 2016, 2017, 2018a,b), or distinct morphemes (e.g., Borer & Ouwayda 2010)
- \$\Rightarrow\$ the underlying syntactic underspecification triggers uniform PF realization despite varied syntactic/semantic behavior, modulo independent differences of the surrounding nominal structures and their morphological realization

2 Facets of functional K and F

2.1 Conceptual gender

- K and F systematically derive female-denoting nouns from MASC nouns
- (1) a. ředitel 'director.M.SG' \rightarrow ředitel-**ka** 'director-**K**:F.SG, a female director' CZECH
 - b. dyrektor 'director.M.SG' \rightarrow dyrektor-**ka** 'director-**K**:F.SG, a female director' POLISH
 - c. kot 'cat.M.SG' \rightarrow kot-ka 'cat-**F:**F.SG, she cat'
 - d. far 'mouse.M.SG' \rightarrow far-a 'mouse-F:F.SG, she mouse'

LA

- e. hamur 'donkey.M.SG \rightarrow hamur-a 'donkey- \mathbf{F} :F.SG., she donkey' HEBREW
- f. daktor 'doctor.M.SG \rightarrow daktor-a 'doctor-F:F.SG., a female doctor' LA
- \Rightarrow *markedness* plays a role:
 - only derivations from unmarked (MASC) to marked (FEM) attested
 - the reversed pattern is non-existent (see, e.g., Pesetsky 2013 for a generalization about the markedness profile of these derivations)
 - for Slavic, no derivation of female-denoting nouns from NEUTER nouns either
 - we take it to mean that the markedness corresponds to switch from [-GENDER], i.e,
 MASC to [+GENDER], i.e., FEM⁵

(2)

Gender Change	F	K
$MASC \Rightarrow FEM$	√	√
$FEM \Rightarrow MASC$	×	×
$* \Rightarrow \text{NEUT}$	n/a	×
$NEUT \Rightarrow *$	n/a	×

⁵We assume that NEUTER is a complex feature, [-PERSON, -GENDER]; see, e.g., Bartošová & Kučerová 2016, 2018.

2.2 Category change

- Slavic **K** systematically functions as a *nominalizer*
- K derives nouns from adjectives
- a. sodová (voda) 'soda.ADJ (water)' → sodov-ka 'soda-K:F.SG, pop'
 b. mielon-y/-a 'minced.ADJ-.masc/fem' → mielon-ka 'luncheon_meat-K:F.SG' POLISH
 - K also derives nouns from verbs
- (4) a. doplnit 'to complement' → dopln-ěk 'complement-K:M.SG, a complement' CZECH
 b. podpalić 'to ignite' → podpał-ka 'accelerant-K:F.SG' POLISH
 - derivation from adjectives and from verbs never generates NEUTER

In Semitic

- we haven't found any instances of a nominalizing **F** but this might be a side-product of the templatic morphology
- consequently, the **F**-like formation might be subsumed within the stem template, i.e., the first spell-out domain

1	^	١
١	J	,

Category Change	F	K
$ADJ \Rightarrow N_{masc}$??	√
$V \Rightarrow N_{masc}$??	√
$\mathrm{AdJ} \Rightarrow \mathrm{N}_{fem}$??	√
$V \Rightarrow N_{fem}$??	√
$Adj \Rightarrow N_{neut}$	n/a	×
$V \Rightarrow N_{neut}$	n/a	×

2.3 Noun to Noun Conversion

- in noun-to-noun conversions,⁶ even though some patterns are more productive than others, all combinations of 3 grammatical genders attested⁷
- the derivation of grammatically FEM nouns from a MASC base is fully productive, and if plausible, often (but not always) ambiguous with a conceptual gender formation
- a. diplomat.M.SG → diplomat-ka 'diplomat-K:F.SG; a briefcase, a female diplomat' CZ
 b. dyplomat.M.SG → dyplomat-ka 'diplomat-K:F.SG; a briefcase, a female diplomat
 POLISH

⁶Czech data are based on Dokulil et al, 1986.

⁷This is not quite right: derivations from neuter are restricted. Adam can tell you more about this at the Q&A period.

- c. stolarz.M.SG 'a carpenter' \rightarrow stolar-ka 'carpenter-**K**:F.SG; carpentry (not a female carpenter)'
- the derivation of grammatically MASC nouns from a FEM base is less productive but still relatively frequent:
- (7) a. kůra 'tree-bark.F.SG' → kor-ek 'bark-K:M.SG, cork'
 b. kora 'tree-bark.F.SG' → kor-ek 'bark-K:M.SG' cork
 POLISH
 - as is the derivation of FEM nouns from a FEM base
- (8) a. kniha 'book.F.SG' → kníž-ka 'book-K:F.SG, a book (less formal)'
 b. košile 'shirt.F.SG' → košil-ka 'shirt-K:F.SG' an undershirt
 - neuter derivational bases are restricted to a handful of nouns
- (9) a. rameno 'shoulder.N.SG → ramín-ko.K:N.SG 'hanger'
 b. biuro 'office.N.SG' → biur-ko 'desk-K:N.SG'
 POLISH
 - as are MASC to MASC conversions
- (10) cukier 'sugar.M.SG/MASS' → cukier-ek 'sugar-**K**:F.SG, a piece of candy' POLISH
 - N-to-N conversions in Semitic are difficult to characterize because of templatic morphology

2.4 Diminutives, their doubles & friends

- DIM formation by **K** is highly productive; no markedness effects
- K displays the gender value of its base noun, (11-a)
- a. jablko 'apple.N.SG' → jablíč-ko 'apple-K:N.SG; a small apple'
 b. jáma 'pit.F.SG' → jam-ka 'pit-.K:F.SG; a small hole'
 c. stůl 'table.M.SG' → stol-ek 'table-.K:M.SG; a small table'
 d. pudło 'box.N.SG' → pudeł-ko 'box.-K:N.SG; a small box'
 e. dziura 'hole.F.SG' → dziur-ka 'hole-K:F.SG; a small hole'
 f. słup 'pole.M.SG' → słup-ek 'pole-K:M.SG; a small pole'
 POLISH
 - the primary diminutive (denoting a small size) derivation in Semitic does not have a separate suffix
 - instead, we see a stem-internal alternation (templatic morphology)
 - as with **K**, this diminutive formation does not change the gender of the base and is productive both for MASC and FEM nominals
- a. da?era 'circle.F.SG → dowerra 'a small circle.F.SG'
 b. bent 'girl..F.SG → bannotta 'a little girl.F.SG'

- c. arnab 'rabbit.M.SG' → arnub 'a small rabbit.M.SG' LA
- d. mHammad.M.SG (proper name) \rightarrow Hammod.M.SG (a familiar 'diminutive' form of the proper name)
- both Slavic and Semitic exhibit double-diminutive formation
- i.e., an additional diminutive-like morpheme is added to a derived diminutive

In Slavic:

- this derivation involves doubling **K**, with the important clarification that only the outside **K** morphologically displays ϕ -features:
- (13) stol-ek 'table- \mathbf{K} :M.SG, a small table' \rightarrow stol-eč-ek 'table- \mathbf{K} :M.SG- \mathbf{K} :M.SG, a very small table' CZ

In Semitic:

- the double-formation combines the stem-internal and the stem-external derivation (e.g., De-Belder et al 2019)
- where the stem-external derivation is realized as **F**
- (14) a. arnab 'rabbit.M.SG' → arnub 'rabbit.DIM.M.SG' → arnub-i 'rabbit.DIM.M.SG-**F**:SG; a cute small rabbit' LA
 - b. mHammad.M.SG (proper name) \rightarrow Hammod..DIM.M.SG' \rightarrow Hammod-i.DIM.M.SG-F:SG; cute/sweet Hammod' LA
 - c. Aya.F.SG (proper name) \rightarrow Ayooš..DIM.F.SG' \rightarrow Ayoosh-i.DIM.F.SG- \mathbf{F} :SG; cute/sweet Aya' LA
 - crucially, the stem-external morpheme **F** is both semantically and syntactically an adjunct (e.g., Wiltschko & Steriopolo 2008 for an argument that some diminutives are structurally adjuncts, while some are functional heads)
 - the most telling piece of evidence is that the stem-external **F**, technically FEM.SG., is invisible for agree
 - the grammatical gender of the double-diminutives is strictly based on the grammatical gender of the nominal base
- (15) al-arnub-i nam b-Hodn-ii the-rabbit.DIM.M.SG-**F**:SG sleep.3M.SG.PST in-lap-my 'the cute small rabbit slept in my lap.'

Doubling yields:

- additional semantic readings \Rightarrow higher degree of a small size, and
- additional pragmatic readings ⇒ affectionate (e.g., Jurafsky 1996, Dressler & Barbaresi 1994, Fassi Fehri 2017)

Augmentatives:

- absent in Slavic⁸
- as with diminutives, the stem-internal derivation yields a pure augmentative interpretation (i.e., a large size of the nominal denoted by the base), and the stem-external derivation either adds a higher degree of large size (or importance etc) or adds additional pragmatic readings
- as with diminutives, the stem-external **F** does not change the grammatical gender of the base nominal ⇒ an adjunct
- raaḥil.M.SG. 'traveler' → raḥḥaal.AUG.M.SG. 'big traveler' → raḥḥaal-**at** raḥḥaal.AUG.M.SG.-**F**:SG 'famous big traveler' MOROCCAN A.; Fassi Fehri 2016: 238, (40)

2.5 Individuation

- F productively individuates (e.g., Zabbal 2002, Acquaviva 2008, Ouwayda 2014, Fassi Fehri 2016, 2018); it forms:
 - singulatives, i.e., individuated batch nouns, and
 - individuated events
- a. Tabšuur 'chalk' (batch noun) → Tabšuur-a 'chalk-F:SG, a piece of chalk'
 b. Saxr 'stone' (batch noun) → Saxr-a 'stone-F:SG; a piece of stone'
- (18) a. raqaṣa raqṣ-an danced dance-ACC 'he danced some dancing'
 - b. raqaṣa raqṣ-at-an/ raqṣ-at-ayn danced dance-**F**:SG-ACC/ dancedance-**F**:SG-DU 'he danced a dance/ two dances'

MOROCCAN A.; Fassi Fehri 2016: 226, (11)

LA

- although certain additional restrictions apply (see, e.g., Zabbal 2002, Borer & Ouwayda 2010), both singulatives and individuated events can be further pluralized
- **F** is also supposed to individuate mass nouns but in this case there always is an additional change in the templatic morphology and a shift in the lexical meaning, indicating an additional structure building and/or additional derivational morphology (similar to specific derivational morphology in Slavic)
- this might not be a pure formation by **F**
- (19) a. sokkar 'sugar.MASS' \rightarrow sokkareyy-i 'sugar-**F**:SG; a sugar bowl'
 - b. foren 'oven.M.SG \rightarrow forneyy-i 'oven-**F**:F.SG; a portable electric oven'

⁸Although there are other lexically specified derivational morphemes that yield related meanings. See, e.g., Steriopolo 2009, 2013, Krizhman 2019).

- K never individuates
- in Slavic, both DIM derived by **K** and its base are systematically ambiguous between mass and count interpretations
- (20) a. Na stole bylo mnoho cukru on table was much sugar.SG/MASS

'There was much sugar on the table.'

- b. Na stole bylo mnoho cukrů
 - on table was many sugar.PL

'There were many kinds of sugar/ pieces of sugar (cubes, packets of sugar) on the table.'

- (21) a. Na stole bylo mnoho cukř-íku
 - on table was much sugar.SG/MASS-K

'There was much cute/sweet/delicious sugar on the table.'

- b. Na stole bylo mnoho cukř-íků
 - on table was many sugar. K:PL

'There were many kinds of adorable sugar/ small pieces of sugar (cubes, packets of sugar) on the table.'

Note:

• the pattern in Slavic and Semitic is consistent with the generalization that DIM individuates mass nouns only if it also changes gender of its base, i.e., it is not an adjunct (e.g., Dutch and German, see, e.g., Borer (2005), p. 92, ft. 6)

2.6 Group formation

- in Semitic, **F** productively derives group formation (Borer 2005, Ouwyada 2014, Kramer & Winchester 2018)
- unlike singulatives, these group denoting nouns cannot be pluralized⁹, suggesting that semantically these are aggregates, i.e., maximized units formed from an already individuated content¹⁰
- (22) mtdyyen 'religious.M.SG, a believer' → mtdyn-i 'religious-F.SG, a religious group' LA
 - in Slavic, group formation by **K** is restricted to numerals, (23-a), quantifiers (Veselovská 2018), (23-b), and to a productive formation of pluralia tantum (Dokulil et al. 1986), (23-c)
 - they are not aggregates, and consequently can be pluralized
- (23) a. dvě děvčata 'two girls' \rightarrow dvoj-ka děvčat 'two-**K**:F.SG girls.GEN, a group of two girls' CZ

⁹The corresponding form exists but it means, for example, a group of female believers, instead of a plurality of groups of believers.

¹⁰In Slavic aggregates of this type are formed by NEUTER (Grimm & Dočekal 2019, Arsenijević 2013). We put neuter formations aside because they are orthogonal to the main question of this talk.

- b. pár děvčat 'a few girls' \rightarrow pár-ek děvčat 'couple-**K**:F.SG girls.GEN, a group of two girls' CZ
- c. běž-ky 'run-**K**:PL, cross-country skis', sjezdov-ky 'go_down-**K**:PL, downhill skis' Cz

3 The case for i^*

The core idea:

- we argue that **K** and **F** are morphological realizations of an underspecified functional head, which we call i^* (Wood & Marantz 2015)
- the functional interpretation of i^* is a function of its structural position $\Rightarrow i^*$ takes its core properties from the head whose features it modifies
- \Rightarrow when i^* attaches to a category defining head, then it functions as a category defining head; when it attaches to an individuating head, then it functions as a an individuation head etc.

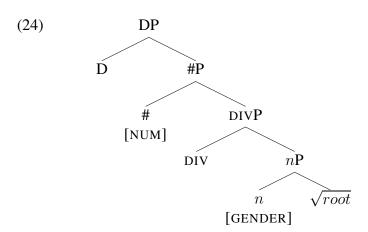
Structural economy:

- we assume that structure building is subject to structural economy
- that is, no feature vacuous structure building is possible
- for i^* to be licensed, the merge of i^* must yield a distinct structure
- by which we mean that the set of features projected after the merge of i^* is not identical to the features in the projection i^* merges with
- since i^* is of the same type as its sister, this economy condition can only be satisfied if the value of the i^* feature is distinct from the value of the corresponding feature in the sister projection

Structural assumptions

- we assume that roots in Slavic (and possibly in Semitic) come with a gender (and a class) index as part of their vocabulary insertion entry (e.g., Acquaviva 2014)
- ullet the root merges with n that carries an unvalued GENDER feature, to be valued by agree with the indexical gender of the root
- nominal roots are by default not individuated (e.g., Borer 2005) ⇒ an individuating projection must be merged (DIVP)
- a counting projection is merged (#P)
- D is merged¹¹

¹¹For the purposes of this talk let's not worry whether or not this D is a phase head. Phasehood as such won't play any role in the data of our concern and the proposal.

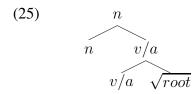


Spell Out

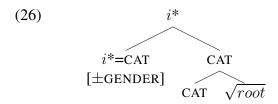
- we follow den Dikken & Dékány (2018) in that syntactic recursion, i.e., the recurrent merge of a head with the same set of syntactic features, is possible only if the two heads are not in the same spell-out domain
- we also assume that in Semitic a morphological template applied to a tri-consonantal root corresponds to a spell out domain

3.1 Category change

- let us demonstrate how i^* works on i^* as a category changing head, i.e., a nominalizer
- in this case, the output of the merge formation is distinct from its input



- in Slavic and Semitic GENDER is tied to nominality and it is the defining feature of n (e.g., Kramer 2015, Veselovská 2019)
- we refine the proposed structure and argue that n in this case is an instantiation of i^* based on a category (CAT) and comprising of a GENDER feature
- since this instantiation of a GENDER feature is a proper syntactic feature, i.e., a feature associated with a functional head, not an indexical feature of a root, this feature is strictly binary ⇒ [±GENDER]



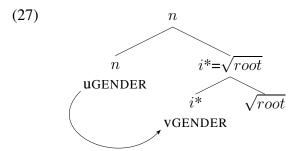
- for (26) to obey structural economy, the projection properties of the new formation must differ from the primary merge
- in this case, the condition is satisfied only if the merge of i^* yields a category change, i.e., only if the original structure didn't contain a [\pm GENDER] feature
- since either value of [±GENDER] will do, we correctly predict that category change yields MASC and FEM

Semitic:

- we argue that the category changing i^* in Semitic is not realized as a separate morpheme
- instead, it triggers an insertion of a distinct template
- structurally, there is no difference between Slavic and Semitic syntax, only an independent difference in morphological realization

3.2 Noun to noun conversion

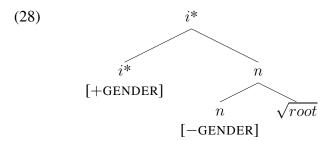
- in noun-to-noun conversion, the category of the root does not change, only gender can change
- ullet we argue that in this case i^* attaches directly to the root, and comprises of a valued GENDER feature
- since i^* merges to a root, it takes root-like properties \Rightarrow a functional root in the sense of Creemers et al. (2019)
- when n is merged, it agrees with the feature of i^* as its closest target
- since this instantiation of a valued GENDER feature is not a proper syntactic feature but instead a property of a functional root, it is indexical and does not have to be binary, and hence a 3-way gender system is attested



3.3 Conceptual Gender

• various evidence points to conceptual gender being introduced by a higher functional head (e.g., Pesetsky 2013, Kramer 2015)

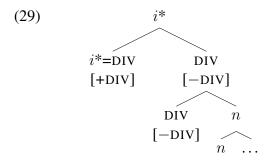
- we argue that in this case the GENDER feature is introduced by an i^* head merged to n, turning i^* into a n-like head
- ullet in order to obey structural economy, the features of the i^* head must be marked relatively to the syntactic features of n
- since n comprises of a GENDER feature, the economy condition is satisfied only if n comes with [-GENDER] and i^* comes with [+GENDER], i.e., it derives a FEM nominal



- crucially, i^* contains the same features as n (albeit differently valued), i.e., it constitutes a recursion of nP
- $\Rightarrow nP$ must be spelled out separately from the i^*
- \Rightarrow in Semitic, the i^* , i.e., **F** is morphologically realized as a suffix, instead of being subsumed within the stem template

3.4 Individuation

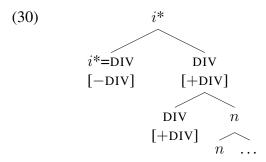
- Semitic has a class of genderless unindividuated nominals, so called batch nouns, in which the individuating functional head, DIV, head is set up as [-DIV]
- when i^* attaches to a DIV projection it becomes a DIV head as well
- ullet for structural economy to be obeyed, i^* must be composed of a [+DIV] feature, yielding the marked individuated interpretation
- this interpretation is absent in Slavic because the equivalent of the DIV projection is optional, and projected only if it is valued as [+DIV] (see, e.g., Dočekal 2012 on kind NPs in Czech, and Déprez 2005, Acquaviva 2018 for an argument that kind nouns lack NumP):



- technically, the merge of i^* constitutes a recursion, i.e., the merge of i^* triggers spell-out of the DIV projection
- ullet in Semitic, the DIV projection is included in the stem template but the i^* is realized as a suffix on the template
- but why does a DIV-like projection morphologically realize as a gender suffix?
- in Semitic, the GENDER feature is privative, i.e., MASC is an absence of a gender feature; gender feature is always FEM
- DIV is home to a classifier feature which is closely related to gender
- ⇒ since Semitic does not have a specific morphology for marking individuation, the DIV feature triggers an insertion of the closest feature with a classifier-like property, namely gender (**F**)

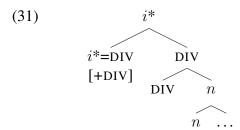
3.5 Group formation

- we argue that the group formation interpretation in Semitic results from the same structure like the individuation structure but the valuation of the DIV feature is reversed
- here the DIV head comes with [+DIV] feature; consequently, for i^* to obey the economy condition, i^* must be valued as [-DIV]
- this derived structure cannot be pluralized because plural requires an individuated, i.e., [-DIV] structure; however, the merge of another layer of i^* is blocked by structural economy because the output of the iterated merge would be equal to the merge before the first i^* was merged
- as with individuation, since there is no designated morphological realization of the DIV feature, morphology realizes the i^* by its closest relative, i.e., gender



In Slavic

- a group interpretation arises only in the context of structures based on quantifiers like 'a couple,' numerals, and pluralia tantum, i.e., roots with [+DIV]-like lexical semantics but without the corresponding individuated structure built in syntax
- we argue that in this structure i^* attaches with a [+DIV] feature to a DIV head that lacks a valued DIV feature



- consequently, the resulting structure is countable
- note that in effect, the derivation of this structure in Slavic is on a par with Semitic singulatives

3.6 Diminutives

- each noun phrase can in principle contain a degree phrase (Morzycki 2009)
- we argue, following much work on Slavic diminutives (e.g., Wiltschko & Steriopolo 2007), that DIM is an adjunct to a DegP

ullet technically, i^* changes the default POS heading DEG to NEG

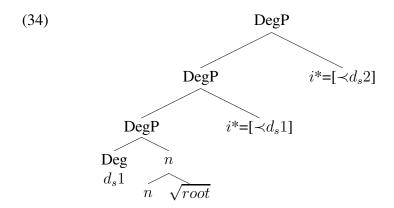
(33) a.
$$[POS] = \lambda g_{\langle e,d \rangle} \lambda x. \text{standard } (g) \leq g(x)$$

b. $[NEG] = \lambda g_{\langle e,d \rangle} \lambda x. \text{standard } (g) \geq g(x)$

- *i** changes the point of reference to be below the standard minimal value
- \bullet gender is transparent since i^* is an adjunct and adjuncts do not project

Doubling

- double even triple diminutives are possible since i^* can apply recursively to reset the scale to the minimal value of its input
- a double DIM formation obeys structural economy only if it yields additional interpretations (Sichel & Wiltschko 2018) \Rightarrow double DIM yields a new degree scale



- when the second instantiation of i^* of Deg merges, we again get recursion, i.e., the structure enforces spell-out
- which gives the right morphology for Semitic ⇒ the primary DIM is stem internal, the second iteration is a suffix

Open question:

• why does i^* surface as **F** in Semitic?

Pragmatic readings

- both single and double DIM not only yield a new degree (based on the bounded nominal projection) but can also be mapped to a pragmatic reading
- pragmatic readings can constitute affection, or derogation (see, e.g., Fontin 2011, Fassi Fehri 2016, 2018)
- these are not a direct product of feature interaction of i^* but rather a mapping of its morphosyntactic effects to the interfaces

4 Conclusions and open questions

- we presented an empirical study that provides evidence that there is a class of structural building operations that are in an important structural sense underspecified
- and that a variety of seemingly varied functional projections can be unified under the underspecification approach
- there are of course many open questions, such as why these particular syntactic interpretations and why they wildly corroborate, and whether we can find *i**-like behavior in other projections as well
- the so-called reflexive in Slavic might be a good candidate for i^* in vP