2024. *Proc Ling Soc Amer* 9(1). 5662. https://doi.org/10.3765/plsa.v9i1.5662.

Subject nominalizations in Setswana

Soo-Hwan Lee*

Abstract. Agent nominalizations of an object-bearing transitive predicate require *of*-insertion in many languages, including English (e.g., *driver* *(*of*) *a truck*). Note that the same transitive predicate is not associated with *of*-insertion in the clausal syntax (e.g., *drive* (**of*) *a truck*). This work provides empirical evidence from Setswana (Bantu) and suggests that *of*-insertion is not necessary in Setswana agent nominalizations and subject nominalizations, more broadly construed. As long as the syntactic licensing requirements are satisfied, they need not resort to *of*-insertion. Additionally, I add weight to the claim that an external argument can be represented inside subject nominalizations (Baker & Vinokurova 2009). Adopting the Phrasal Layering approach (Alexiadou & Schäfer 2010, among others), I argue that parallels can be drawn between the nominal domain and the clausal domain in Setswana.
Keywords. agent/subject nominalizations; *of*-insertion; allomorphy; Setswana

1. Introduction. It is commonly assumed that an object-bearing transitive predicate inside an agentive nominal requires *of*-insertion. English is no exception, as shown in (1).

(1) a. a driver **of** a truck
b. *a driver a truck

Of-insertion, however, does not carry over to the clausal syntax in the same manner. That is, a transitive predicate is realized with an object without *of*-insertion, as shown in (2). (2) showcases the opposite pattern observed in (1).

(2) a. *He drives **of** a truck.
b. He drives a truck.

Discrepancies such as the one observed between (1) and (2) have led to the idea that the nominal syntax and the clausal syntax are not the same. This work provides empirical evidence from Setswana (Bantu) suggesting that *of*-insertion is not a must in the nominal domain. (3) shows that the nominal domain and the clausal syntax display a fair amount of transparency.

(3) a. mo-kgweets-a di-teraka
1-drive-A 10-truck
'a driver of trucks'

b. o-kgweets-a di-teraka
3.SG-drive-A 10-truck
'He drives trucks.'

* I express my gratitude to Seabelo C. John, my native consultant, for providing me the opportunity to work on Setswana. I also thank Chris Collins, Alec Marantz, Richie Kayne, Gary Thoms, Vicki Carstens, Claire Halpert, Asia Pietraszko, Ruth Kramer, Jim Wood, Maria Kouneli, and Selikem Gotah for their helpful comments. All remaining errors are my own. Author: Soo-Hwan Lee, New York University (soohwan.lee@nyu.edu).

It is worth noting, however, that *of*-insertion is necessary when the nominal suffix is realized as *-i* instead of *-a*. Contrast (3) and (4).

(4) a. mo-kgweets-**i** **wa** di-teraka
1-drive-**I** **1.of** 10-truck
'a driver of trucks'

b. *mo-kgweets-**i** di-teraka
1-drive-**I** 10-truck
Intended: 'a driver of trucks'

I argue that the discrepancy between (3a) and (4a) is attributed to the different nominal licensing mechanisms adopted in the language. In (3a), v licenses the object *di-teraka* 'trucks', and as a result v is realized as *-a* in PF. In (4a), v cannot license the object, since the object is already licensed by P (*wa*). Due to the licensing failure, v is realized as *-i*. In this regard, the presence or the absence of *of*-insertion is associated with v's ability to license inside the nominal domain.

Additionally, I argue alongside Baker & Vinokurova (2009) that the argument structure established inside subject nominalizations can accommodate an external argument (EA). The evidence comes from reflexive anaphor binding:

(5) a. mo-**i**-pola-i
1-**REFL**-kill-I
'one who kills himself/herself'

b. mo-**i**-that-i
1-**REFL**-love-I
'one who loves himself/herself'

(5) suggests that an antecedent argument needs to be introduced in order for the incorporated reflexive argument *i-* to be syntactically bound (see Mchombo 2004). Following Baker & Vinokurova (2009), I posit that PRO is the antecedent introduced in Spec,vP and that this EA binds the reflexive anaphor. Overall, this work provides empirical evidence from Setswana in suggesting that *of*-insertion is not necessary and that an EA can be accommodated inside subject nominalizations. Hence, the nominal domain aligns with the clausal domain in syntax.

The organization of this work is as follows: Section 2 lays out the basic facts about Setswana nominals and the empirical data on Setswana subject nominalizations. Section 3 introduces the analysis of the paper. Section 4 presents cross-linguistic data and the theoretical implications. Section 5 concludes.

2. Setswana nominals. In this section, I present the basic data on Setswana simple nouns and how agent/subject nominalizations are derived in the language.

2.1. BASIC DATA. Similar to other Bantu languages, Setswana displays a number of noun class prefixes, as shown in Table 1.

Noun classes 1 and 2 are used for human-denoting nominals as shown in (6). Noun class 1 is associated with singular entities, and noun class 2 is associated with plural entities.

Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
mo-	*ba-*	*mo-*	*me-*	*le-*	*ma-*	*se-*	*di-*
Class 9	Class 10	Class 11	Class 12	...	Class 16	Class 17	Class 18
ø	*di-*	*lo-*	*di-*	...	*fa-*	*go-*	*mo-*

Table 1. Noun classes in Setswana (Cole 1975:69)

(6) a. **mo**-/**ba**-tho
1-/**2**-person
'a person / people'

b. **Mo**-/**Ba**-tšwana
1-/**2**-Tswana
'a Tswana tribesman / Tswana tribesmen'

In many cases, agentive nominals host noun classes 1 and 2. Agentive nominals are also realized with a suffix. In (7), the suffix *-i* is realized together with a noun class prefix. In fact, the templates '1-V-i' and '2-V-i' are productively used in Bantu agent nominalizations (see Du Plessis 1997 an Mletshe 2010).[1]

(7) a. **mo**-/**ba**-bin-**i**
1-/**2**-dance-**I**
'a dancer/ dancers'

b. **mo**-/**ba**-tlod-**i**
1-/**2**-jump-**I**
'a jumper / jumpers'

The pattern observed in (7) extends to subject nominalizations. More discussion on subject nominalizations is provided in the following subsection.

2.2. SUBJECT NOMINALIZATIONS. Unaccusative predicates can surface together with a noun class and the suffix *-i*, as shown in (8).

(8) a. mo-/ba-**sw**-i
1-/2-**die**-I
'one who dies (dier) / those who die (diers)'

b. mo-/ba-**w**-i
1-/2-**fall**-I
'one who falls (faller) / those who fall (fallers)'

[1] Noun classes 7 and 8 can also be used to derive agentive nominals. In such cases, profession is added to their meaning: ***se-/di-****bin-i* (**7**-/**8**-dance-I, 'a professional dancer / professional dancers').

This calls for a reconceptualization of the type of nominalizations under discussion. Baker & Vinokurova (2009) refer to it as 'subject nominalization' based on their observation on a similar pattern observed in Gĩkũyũ (Bantu):[2]

(9) a. mu-ku-i
1-die-I
'one who dies'

b. mu-twek-i
1-melt-I
'one who melts' (Gĩkũyũ, Baker & Vinokurova 2009:547)

Transitive predicates can be realized with an overt object DP in subject nominalizations. (10) shows that the verbs are nominalized in the presence of the associative marker *wa* 'of.' As indicated in (10), *of*-insertion is obligatory in the presence of the suffix *-i*.

(10) a. mo-kgweets-**i** *(**wa**) teraka
1-drive-**I** **1.of** 9.truck
'a driver of a truck'

b. mo-palam-**i** *(**wa**) thaba
1-climb-**I** **1.of** 9.mountain
'a climber of a mountain'

A transitive predicate can be nominalized in the absence of *wa* 'of.' (11) shows that *of*-insertion must be absent in cases where the suffix *-a* surfaces instead of *-i*.

(11) a. mo-kgweets-**a** (***wa**) teraka
1-drive-**A** **1.of** 9.truck
'a driver of a truck'

b. mo-palam-**a** (***wa**) thaba
1-climb-A 1.of 9.mountain
'a climber of a mountain'

Unlike English compounds (e.g., *a truck-driver*), a full-fledged object DP can participate in Setswana subject nominalizations, as shown in (12). Inside the object DP, a possessive pronoun is realized in (12a), a demonstrative is realized in (12b), and a relative clause (RC) is realized in (12c).

[2] It is worth mentioning that agent nominalizations are a proper subset of subject nominalizations. Agent nominalizations denote agentive entities while subject nominalizations denote agentive or non-agentive entities. Hence, they are not mutually exclusive.

(12) a. mo-kgweets-a [**teraka yame**]
1-drive-A **9.truck my**
'a driver of my truck'

b. mo-kgweets-a [**di-teraka tse**]
1-drive-A **10-truck these**
'a driver of these trucks'

c. mo-kgweets-a [**di-teraka tse yo Kitso amoratang**]
1-drive-A **10-truck these REL Kitso likes**
'a driver of these trucks that Kitso likes'

Note that (12) cannot be replicated using English compounds due to the deprived size of their object DP. This is demonstrated in (13).

(13) a. *a [**my truck**]-driver

b. *a [**these trucks**]-driver

c. *a [**these trucks that John likes**]-driver

The empirical picture sketched out so far suggests that *of*-insertion is not necessary in Setswana subject nominalizations. A question arises as to why this is the case. In the next section, I provide an analysis based on nominal licensing.

Before moving on, I would like to address one other issue that will contribute to the overall discussion. Baker & Vinokurova (2009) argue that anaphor binding is possible inside Gĩkũyũ subject nominalizations. The reflexive marker *ĩ-* participates in the derivations, as shown in (14).

(14) a. mũ-**ĩ**-end-i
1-**REFL**-like-I
'one who likes himself/herself'

b. mũ-**ĩ**-yamb-i
1-**REFL**-pride-I
'one who is full of himself/herself' (Gĩkũyũ, Baker & Vinokurova 2009:548)

Adopting Mchombo's (2004) analysis that the reflexive marker is a cliticized pronoun (anaphor), Baker & Vinokurova (2009) claim that PRO is introduced in the derivation as the binder (antecedent). Here, PRO is the EA that undergoes external merge (EM) in Spec,vP. This way of viewing the syntax of subject nominalizations accounts for the following Setswana data in (15), which bears resemblance with (14):

(15) a. mo-**i**-pola-i
1-**REFL**-kill-I
'one who kills himself/herself'

b. mo-**i**-that-i
1-**REFL**-love-I
'one who loves him/herself'

The reciprocal marker *-an* is also realized in Setswana subject nominalizations. Since reciprocals refer to a plural entity, noun class 2 has to be employed instead of noun class 1 (see Section 2.1). (16) illustrates this point.

(16) a. ba-tsa-**an**-i
2-marry-**RECP**-I
'those who marry each other'

b. *mo-tsa-**an**-i
1-marry-**RECP**-I
Intended: 'those who marry each other'

Now that we have all the pieces of evidence, I will move forward and present an analysis on the patterns displayed in Setswana subject nominalizations.

3. The analysis. The (im)possible patterns for Setswana subject nominalizations are summarized in (17). A generalization that emerges from these patterns is that the suffix *-i* is realized when there is no object DP or when the DP is accompanied by *of*-insertion. If this requirement is not satisfied, the suffix *-a* is realized instead.

(17) a. 1-/2-V-i
b. *1-/*2-V-a
c. 1-/2-V-i OF DP
d. *1-/*2-V-a OF DP
e. *1-/*2-V-i DP
f. 1-/2-V-a OF DP
g. 1-/2-i-V-i
h. *1-/*2-i-V-a
i. 1-/2-V-an-i
j. *1-/*2-V-an-a

The analysis of this work is based on nominal licensing. I assume that all overt DPs must be licensed in syntax. (18) is subject to an updated version of the Activity Condition (Chomsky 2001):

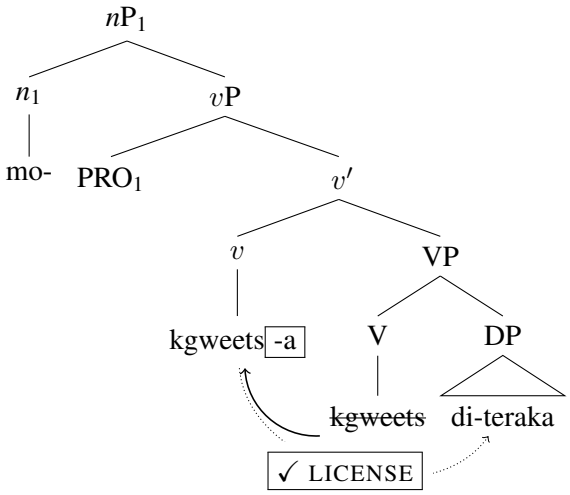
(18) Nominal licensing condition
All overt DPs must be syntactically licensed.

When the preposition (P) 'of' is present in the derivation, P licenses the object DP. When P is absent, however, Setswana relies on another head that licenses the DP. I argue that this head is v, as illustrated in (19). When v licenses the object DP, v is realized as *-a* as an indication of successful licensing. When P licenses the DP instead of v, v is realized as *-i* as a reflex of its failure to license, as shown in (20). Note that P is structurally closer to the object DP than v is. Hence,

v's licensing failure follows from structural locality. Both (19) and (20) host PRO as their EA in Spec,vP. This accounts for the anaphor binding facts presented in Section 2.2. Following Baker & Vinokurova (2009), I posit that PRO is coindexed with the nominalizer (n) and the entire nP. The noun class prefix is in n (see Lee & Lee 2019; Fuchs & van der Wal 2022). In order to capture the correct morpheme order, V-to-v movement applies in the derivation. I follow Baker's (1985) Mirror Principle.

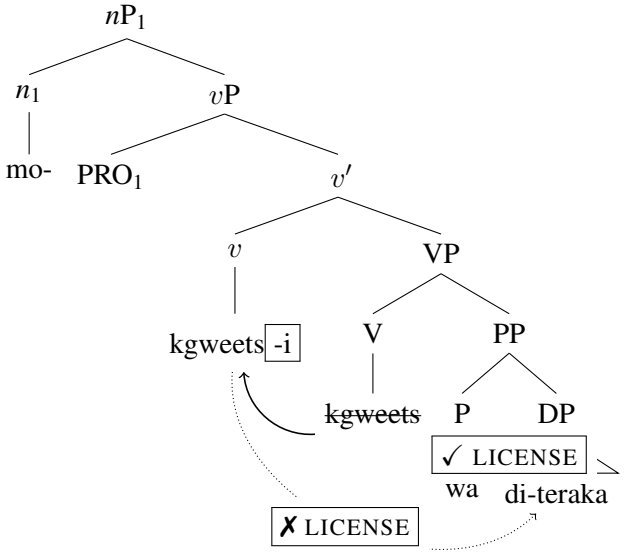
(19) a. mo-kgweets-**a** di-teraka
1-drive-**a** 10-truck
'a driver of trucks'

b.



(20) a. mo-kgweets-**i** wa di-teraka
1-drive-**I** 1.of 10-truck
'a driver of trucks'

b.



Here, it is worth mentioning an alternative analysis that comes close to handling the *-i*∼*-a* alternation. It runs on a couple of phonological assumptions. One may assume that the glide *w* that linearly intervenes between the suffix *-i* and *a* of *wa* is phonologically reduced to a point where vowel hiatus results between *-i* and *a*. In order to preserve the phonological well-formedness, *-i* is deleted.[3] At this point, it is necessary to adjudicate which of the two competing hypotheses provided so far best suits the overall empirical picture. A piece of evidence that works in favor of the syntactic approach as opposed to the phonological approach is presented in (21). In (21), the associative marker 'of' undergoes agreement/concord with noun class 2 and is spelled out as *ba* instead of *wa*.

(21) ba-kgweets-i **ba** di-teraka
2-drive-I **2.of** 10-truck
'drivers of trucks'

It is difficult to make the case that the bilabial stop *b* undergoes phonological reduction in the same way as the glide *w* does. The bilabial stop is a full consonant and does not carry the defining characteristics of a vowel. Hence, vowel hiatus between *-i* and *a* cannot be at play. Nevertheless, the realization of the suffix *-a* in the absence of the associative marker is possible, as shown in (22). The syntactic analysis put forward in (19) and (20) works just as well for (21) and (22).

(22) ba-kgweets-**a** di-teraka
2-drive-**A** 10-truck
'drivers of trucks'

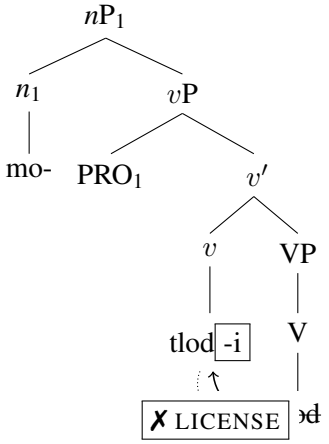
Recall that PRO is introduced in subject nominalizations. While PRO can be introduced as an EA, as shown in (23), PRO can also be introduced as an internal argument (IA) as long as it ends up as the subject coindexed with the entire noun phrase (see Baker & Vinokurova 2009). In (24), for instance, PRO is realized as the IA of an unaccusative predicate. Since there are no other arguments introduced in the derivation, PRO is the subject. This is similar to how an IA can be the subject of an unaccusative predicate in the clausal domain.

Moreover, subject nominalizations bearing unergative predicates are analyzed on a par with those bearing unaccusative predicates with respect to the *-i*∼*-a* alternation. Based on (17a) and (17b), the suffix *-i* surfaces together with intransitive (unergative and unaccusative) predicates, but the suffix *-a* does not. I argue that this is because the only argument introduced in this type of derivation, namely PRO, is exempt from the nominal licensing condition outlined in (18). This is also in accordance with a line of research that suggests PRO cannot be Case-licensed in syntax.

[3] I thank Ruth Kramer (p.c.) for pointing this out.

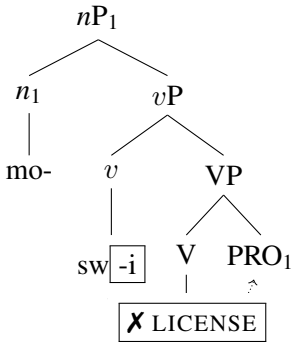
(23) a. mo-tlod-i
1-jump-I
'a jumper'

b.



(24) a. mo-sw-i
1-die-I
'one who died (dier)'

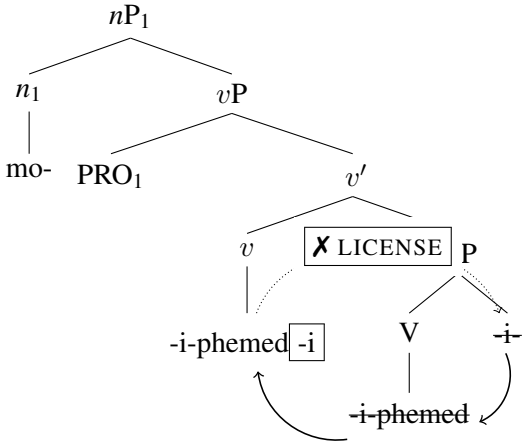
b.



Now we turn to the anaphor binding cases. While the suffix *-i* is compatible with subject nominalizations showcasing anaphor binding, the suffix *-a* is not. Adopting Mchombo (2004), I take the reflexive marker in Bantu to be an incorporated nominal. Nominals that undergo incorporation are syntactically reduced in size, implying that they are less than a full-fledged DP. Based on (18), then, v does not license the reflexive marker, since it is not a DP. Consequently, v is realized as *-i*, not *-a*. (25) fleshes out the details to the derivation.

(25) a. mo-i-phemed-i
1-REFL-defend-I
'one who defends himself/herself'

b.



Based on the discussion so far, the suffixes *-i* and *-a* can be treated as allomorphs conditioned by the presence or the absence of nominal licensing. Adopting Distributed Morphology (Halle & Marantz 1993), I provide the following vocabulary insertion (VI) rules for the two suffixes:

(26) VI rules for the suffixes *-i* and *-a*

a. $v \leftrightarrow$ *-i* / [$_{nP}$ *n* [$_{vP}$ — ✗ LICENSE (DP)]]

b. $v \leftrightarrow$ *-a* / elsewhere

In the next section, I show that subject nominalizations are attested in languages from typologically distinct families and that they are not necessarily rare cross-linguistically (contra Baker & Vinokurova 2009).

4. Cross-linguistic data and theoretical implications. In addition to Gĩkũyũ and Setswana, Oshiwambo (Bantu) hosts subject nominalizations (Lee & Ndapo, In progress). Oshiwambo allows unaccusative predicates inside subject nominalizations. In (27), the verbs *gw* 'to fall' and *s* 'to die' are associated with the suffix *-i*, denoting subject nominalizations. The verbs are also realized with an augment (AUG) and a noun class, as is usually the case with Oshiwambo noun phrases in general.

(27) a. o-mu-**gw**-i
AUG-1-**fall**-I
'one who falls (faller)'

b. o-mu-**s**-i
AUG-1-**die**-I
'one who dies (dier)'

The reflexive marker *i-* and the reciprocal marker *than* can be realized inside Oshiwambo subject nominalizations, as shown in (28). This is similar to how Gĩkũyũ and Setswana allow anaphor binding inside subject nominalizations.

(28) a. o-mwi-**i**-dheng-i
AUG-1-**REFL**-hit-I
'one who hits himself/herself'

b. a-a-hokanna-**than**-i
AUG-2-marry-**RECP**-I
'those who marry each other'

Ewe (Tongugbe dialect, Kwa) also showcases subject nominations. Gotah & Lee (2024) report that unaccusative predicates can take part in Ewe subject nominalizations:

(29) a. Kofi ne yi kɔdzí kaba-e anyené mé-vá zu **ku-ku-lá** ɔ́.
Kofi 3SG go hospital early-TOP SUBJ NEG-come be **die-die-LA** NEG
'Had Kofi gone to the hospital earlier, he wouldn't have (been a dier) died.'

b. Ati-ɛ́ wó ŋe-ŋe ná bé Kofi vá zu **ge-ge-lá**.
stick-DEF POSS.SG break-break give COMP Kofi come be **fall-fall-LA**
'The fact that the stick broke is the reason Kofi (became a faller) fell.'

Similar to the languages we have covered so far, Ewe allows anaphor binding inside subject nominalizations:

(30) a. [**Đokoé**-dzí-ɖu-lá] Kofi nyɔ́.
REFL-top-win-LA Kofi be
'Kofi is a disciplined person.'

b. Kofi kplí Adzo wó-vá zu [**wónɔɛɔ**-ɖe-lá-ɔ́].
Kofi and Adzo 3PL-come be **RECP**-marry-LA-PL
'Kofi and Adzo become a married couple.'

Wá·šiw (Isolate, USA) patterns together with Gĩkũyũ, Setswana, Oshiwambo, and Ewe in that the language hosts subject nominations. Hanink (2021) reports that Wá·šiw unaccusative predicates can take part in subject nominalizations, as shown in (31).

(31) da-**góta?**
3.UN-**break**
'something that is broken' (Wá·šiw, Hanink 2021:12)

Hanink (2021) demonstrates that anaphor binding and accusative (ACC)-licensing are also possible inside these derived nominals. (32) illustrates these points.

(32) a. Ramona de-**gum**-díʔye? L-éʔ-i
Ramona 3.UN-**REFL**-call 1-be-IND
'My name is Ramona (Lit. 'one who calls herself Ramona')'

b. **t'ánu** t'-íšiw-ha
person.ACC 3.UN-get.well-CAUS
'person healer (Lit. 'one who heals people')' (Wá·šiw, Hanink 2021:12)

Subject nominalizations in all of the languages discussed so far are best analyzed under the Phrasal Layering analysis (Alexiadou & Schäfer 2010, among others). The Phrasal Layering analysis highlights the similarities between the nominal and clausal syntax. Hence, verbal projections, such as VP and *v*P, can be employed for nominalizations under this approach. It follows, then, that EAs and IAs can be introduced in nominalizations. This accounts for the realization of reflexive markers, nominal licensing, and the distinction between unergative and unaccusative predicates.

5. Conclusion. This work has provided a way of obviating *of*-insertion in Setswana subject nominalizations. I have also demonstrated that argument structure can be established in this domain. Parallels have been drawn between the nominal and clausal syntax. I have argued that the Phrasal Layering approach (Alexiadou & Schäfer 2010) offers a way of accounting for the overall data. Further, I have shown that subject nominalizations may not be as rare as what has been assumed in the previous literature (Baker & Vinokurova 2009). In addition to Gĩkũyũ and Setswana, evidence from Oshiwambo, Ewe, and Wá·šiw has been presented to demonstrate this point.

References

Alexiadou, Artemis & Florian Schäfer. 2010. On the syntax of episodic vs. dispositional *-er* nominals. In Artemis Alexiadou and Monika Rathert (eds.), *The syntax of nominalizations across languages and frameworks*, 9–39. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110245875.9.

Baker, Mark. 1985. The mirror principle and morphosyntactic explanation. *Linguistic Inquiry* 16(3). 373–415. http://www.jstor.org/stable/4178442.

Baker, Mark & Nadya Vinokurova. 2009. On agent nominalizations and why they are not like event nominalizations. *Language* 85(3) 517–556. http://www.jstor.org/stable/40492892.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language,* 1–52. Cambridge, MA: MIT Press.

Cole, Desmond. 1975. *An introduction to Tswana grammar*. Cape Town: Longman.

Du Plessis, Jacobus Albertus. 1997. *Morphology of the African languages: Imofoloji Yeelwimi Zaseafrika*. Stellenbosch, ZA: University of Stellenbosch, Department of African Languages.

Fuchs, Zuzanna & Jenneke van der Wal. 2022. The locus of parametric variation in Bantu gender and nominal derivation. *Linguistic Variation* 22(2). 268–324. https://doi.org/mkzn.

Gotah, Selikem & Soo-Hwan Lee. 2024. Syntactic negation in Ewe (Tongugbe) agent nominalizations. *Syntax* (online first) https://doi.org/10.1111/synt.12280.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Ken Hale and Samuel J. Keyser (eds.), *The view from building 20: Essays in honor of Sylvain Bromberger*, 111–176. Cambridge, MA: MIT Press.

Hanink, Emily A. 2021. Restructuring and nominalization size. In Sabine Laszakovits & Zheng Shen (eds.), *The size of things I: Structure building*, 3–23. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.5524274.

Lee, Soo-Hwan & Doo-Won Lee. 2019. Nominal mismatches in Swahili locatives. *Proceedings of the Linguistic Society of America (PLSA)* 4(5). 1–11. https://doi.org/mkzr.

Mchombo, Sam. 2004. *The syntax of Chichewa*. Cambridge: Cambridge University Press.

Mletshe, Loyiso Kevin. 2010. *Deverbal nominals in Xhosa*. Stellenbosch, ZA: University of Stellenbosch dissertation.