Chapter 1

On the expression of information structure in Bantu

Jenneke van der Wal[a], Allen Asiimwe[b], Patrick N. Kanampiu[c], Elisabeth J. Kerr[d], Zhen Li[e], Amani Lusekelo[f], Nelsa Nhantumbo[g] & Ernest Nshemezimana[h]

[a]Leiden University [b]Makerere University [c]University of Edinburgh [d]Ghent University [e]Peking University [f]Dar es Salaam College of Education [g]Universidade Eduardo Mondlane [h]Université du Burundi

This chapter provides an introduction to the volume. It sketches the background to the BaSIS project and methodology, explains the basics of information structure and terms in the field, exemplifies common constructions for information structure in Bantu languages (specifically cleft constructions, word order, and predicate doubling), and concisely explains the main diagnostics from the BaSIS methodology that are used in the following chapters.

1 Introduction

All users of all languages structure the information they give to their addressees so that the addressees can more easily incorporate the information into their current knowledge. This process involves indicating what is already known and what is new, and also which information contrasts with what the addressee might have in mind. Languages differ in terms of which linguistic strategies they have available for the language user to employ to express information structure. While Germanic languages for example use pitch accent to express what information is new or contrastive ("She saw the MAN" vs "She SAW the man"), in the Bantu (Niger-Congo) languages, morphosyntactic strategies are key for the expression of information structure (Downing & Hyman 2016, van der Wal 2015, Downing &

Marten 2019), with many different strategies available. For example, we can see in example (1) how word order reflects information structure in Kirundi. Here, the constituent in clause-final position is in focus: this final element forms the new or contrastive information (a more precise definition will follow in Section 3).

(1) Kirundi (Chapter 5, Nshemezimana & van der Wal 2025 [this volume])[1]

   a. (Where did the mother (that we were talking about) give the milk to the children?)
      Nya muvyéeyi yaheereye amatá abáana **mu nzu**.

nya	mu-vyéeyi	a-a-∅-há-ir-ye
1.DEM7	1-mother	1SM-N.PST-CJ-give-APPL-PFV

a-ma-tá
AUG-6-milk

a-ba-áana	**mu n-zu**
AUG-2-child **18**	**9-house**

      'It is [in the house]_FOC that the mother gave milk to the children.'

   b. (What did the mother (that we were talking about) give to the children in the house?)
      Nya muvyéeyi yaheereye abáana mu nzu **amatá**.

nya	mu-vyéeyi	a-a-∅-há-ir-ye
1.DEM7	1-mother	1SM-N.PST-CJ-give-APPL-PFV

a-ba-áana	mu
AUG-2-child	18

n-zu	**a-ma-tá**
9-house	**AUG-6-milk**

      'It is [the milk]_FOC that the mother gave to the children in the house.'

   c. (To whom did the mother (that we were talking about) give the milk in the house?)
      Nya muvyéeyi yaheereye amatá mu nzu **abáana**.

nya	mu-vyéeyi	a-a-∅-há-ir-ye
1.DEM7	1-mother	1SM-N.PST-CJ-give-APPL-PFV

a-ma-tá	mu
AUG-6-milk	18

n-zu	**a-ba-áana**
9-house	**AUG-2-child**

      'It is [to the children]_FOC that the mother gave milk in the house.'

Another example of morphosyntactic expression of information structure in Bantu is how focus determines the form of the verb in Makhuwa-Enahara. In (2), we see that the so-called conjoint verb form (CJ) is only felicitous in a context that indicates focus on the constituent following the verb, whereas the disjoint verb

---

[1]See Section 6 and the abbreviation list at the end of this chapter for information on the representation of examples and glosses.

form (DJ) is not acceptable in such an environment (and instead is used when the focus is on the verb itself).

(2) Makhuwa-Enahara (Kerr & van der Wal 2023: 463)

    CJ #Context 1: Are you frying or grilling the fish? (focus on verb)
        Context 2: What are you grilling? (focus on object)
        Ki**n**aánéélá ehopá.
        ki-**n**-aaneel-a   ehopa
        1SG.SM-PRS.CJ-grill-FV 9.fish

        'I'm grilling a fish.'

    DJ Context 1: Are you frying or grilling the fish? (focus on verb)
        #Context 2: What are you grilling? (focus on object)
        Ki**ná**mwáaneéla (ehópa).
        ki-**na**-aaneel-a   ehopa
        1SG.SM-PRS.DJ-grill-FV 9.fish

        'I'm grilling it (the fish).'

Given these examples, an interesting question arises: how exactly does information structure influence the syntax of Bantu languages? This is the question underlying the project "Bantu Syntax and Information Structure" (BaSIS), funded by an NWO Vidi grant and hosted at Leiden University from December 2017 to July 2023. Insights into this question are important in order to develop linguistic models that accurately reflect how information structure interacts with other components of the grammar. In order to answer this more theoretical question, however, it is first necessary to better understand how information structure works cross-linguistically. Hence, the first step on the path to a better model of information structure in the grammar is to find systematic data on the expression of information structure in Bantu languages. That descriptive endeavour is the aim of this book.

This book presents detailed descriptions of the expression of information structure in the eight Bantu languages that formed part of the BaSIS project: Tunen (Cameroon), Teke-Kukuya (Congo), Kîîtharaka (Kenya), Kirundi (Burundi), Rukiga (Uganda), Kinyakyusa (Tanzania), Makhuwa-Enahara (Mozambique), and Cicopi (Mozambique). While a more in-depth analysis is given for some topics in some languages, the consistent core aim is to describe the morphosyntactic strategies that are used in each language to structure information. The book is therefore aimed at Bantuists, typologists, and linguists interested in information structure.

The current chapter is intended as an introduction to the investigation of the expression of information structure in Bantu, serving as common background to the chapters on individual languages. We introduce the languages of the BaSIS project in Section 2, discuss the key conceptual notions in information structure as well as the diagnostics for information structure used in the project in Section 3, and mention three common morphosyntactic strategies for the expression of information structure in Bantu in Section 4: cleft constructions, word order, and predicate doubling. We then describe the general methodology used in the project in Sections 5 and 6 and finally provide a reading guide of what (not) to expect in the book in Section 7.

2 The languages in the book

As mentioned above, the Bantu languages are interesting to study for their information structure, as it seems to affect the core of the grammar, with visible effect on morphology and syntax. Another reason that such research is interesting for comparative purposes is the fact that the Bantu language family is large, comprising of some 555 closely-related languages (Hammarström 2019) spoken across a large geographical range of sub-Saharan Africa. Despite the number of languages and geographic spread, these languages show a lot of similarities: all languages have a noun class system where the class of a noun is visible in the morphology, on the noun itself and/or in the concord in the phrase and agreement in the clause; all languages apart from those in the north-west show extensive agglutinative verbal morphology, with the verb taking derivational and inflectional suffixes and prefixes; and a large majority of categories are head-initial in a large majority of languages (see for overviews of Bantu clausal and nominal syntax Downing & Marten 2019 and Van de Velde 2019, respectively). This relative grammatical similarity makes the Bantu languages important for the discovery of general linguistic principles, as it creates the possibility to study microvariation in the expression of information structure, giving us potential insight into the parameters of language variation as well as language change.

One of the strengths of the BaSIS project is its collaboration between native speaker linguists and external linguists. In this way, the investigation benefits from the inside as well as the outside perspective, with each researcher asking questions and making associations that the other would not as easily get to. The selection of languages for the project was therefore dependent on the collaborations, plus the two languages investigated by the PhD candidates, leading to what one might call a convenience sample (only that ten languages is perhaps

too small to even properly be called a sample). Although there is a fair geographical spread, the languages are very few and not at all a representative sample of the Bantu family, with a greater representation of Eastern Bantu languages. We therefore hope that this research may inspire others to explore the same questions for other (Bantu) languages in order to arrive at a more complete understanding of information structure in the family.

The languages at the start of the project were the following, each specified with its ISO code, Guthrie number,[2] country where it is primarily spoken, and the respective primary researcher indicated (see also the map in Figure 1, page 6):

- Tunen (ISO 639-3 [tvu], Guthrie A44, Cameroon; Elisabeth J. Kerr)

- Teke-Kukuya ([kkw], B77a, Congo; Zhen Li)

- Kîîtharaka ([thk], E54, Kenya; Patrick N. Kanampiu)

- Kirundi ([run], JD62, Burundi; Ernest Nshemezimana)

- Rukiga ([nyn], JE14, Uganda; Allen Asiimwe)

- Kinyakyusa ([nyy], M31, Tanzania; Amani Lusekelo)

- Makhuwa-Enahara ([vmw], P31E, Mozambique; Jenneke van der Wal)

- Otjiherero ([her], R30, Namibia; Jekura U. Kavari)

- Changana ([tso], S53, South Africa, Mozambique; Aurélio Z. Simango)

- Cicopi ([cce], S61, Mozambique; Nelsa Nhantumbo)

The map below shows the approximate locations of these languages, based on coordinates from Glottolog (Hammarström et al. 2022). Further information about each of the languages, such as geography and demography, is given in the respective chapters.

Two of these languages unfortunately do not form part of this book, as you will have noticed. The COVID-19 pandemic caused significant delay in the fieldwork, which meant that the collaboration with Jekura Kavari planned for 2020/21 was no longer feasible. Otjiherero was therefore removed from the project. For

---

[2]As is customary in Bantu linguistics, an alphanumeric classification is provided based on the classification of Guthrie (1948) and updated by Maho (2003, 2009) and Hammarström (2019). This Guthrie classification is based on the location of the language spoken, ranging from zone A (languages in the north-west) to zone S (languages in southern Africa).

Maps © Thunderforest, Data © OpenStreetMap contributors

Figure 1: Map showing the geographical distribution of the languages in the BaSIS project after the removal of Herero.

Changana, the collaboration started with a fieldwork visit in 2019 which resulted in a database, but, unfortunately, the collaborator for the language, Aurélio Simango, passed away in January 2022, hence the work was not completed.

For the eight remaining languages, each chapter in this book describes the morphosyntactic strategies that express (some aspect of) information structure. We do not give the background on general properties of the grammar of each language, instead diving right into the expression of information structure. For some languages, the description is more in depth than for others, as for some we could draw on already published sources (e.g. Nshemezimana's 2016 thesis on Kirundi information structure) or separate articles published during the project (e.g. Li 2020 on the Kukuya passive, and Asiimwe & van der Wal 2021 on the Rukiga contrastive topic marker). While some linguistic strategies are covered in every chapter, such as word order and clefts, each language also has its own specific topics, depending on the special features of the language, and also on the interest of the researcher(s) involved. For the data collection in all languages we used the same methodology, which we explain in detail in Sections 4 and 6. First, however, we explain the concepts related to information structure that form the research focus of the chapters in this book.

3 Information structure

There are two main aspects that play a role in information structure. The first is the *activation status* of referents, and the second is the *functions* that referents may take within the sentence (specifically, topic and focus). We discuss each of these in turn.

3.1 Referent activation status

As introduced above, information structure looks at how we construct sentences dependent on how we want to package information for the addressee. So, what counts as information? One key component is (the mental representation of) referents, and the expressions used to refer to them. For example, we may start a story by introducing the main characters (e.g. "There was a chicken and a cockroach"), and then continue to refer to these characters over the story ("the cockroach", "he", "that chicken" etc.). Here, there is an important distinction between on the one hand the *referents* themselves – which can either be conceptualised as objects existing in the real world, or as mental representations existing in the minds of the speakers – and on the other hand the linguistic *expressions* used

to refer to those referents. In this way, a single referent (e.g. the chicken from the story above) may be referred to using different linguistic strategies (e.g. a noun phrase "the chicken", pronoun "it", or modified noun phrase "that young chicken"). *Referent tracking* looks at how such referencing works in discourse.

Referents have a certain *activation status*, relating to their cognitive accessibility. If they have not been mentioned and are not in the direct context, they are *inactive*, whereas if they are currently under discussion, they are highly *active* (see e.g. Chafe 1987). A referent's activation status has an influence on which expressions we use to refer to it, as proposed in various guises by Givón (1976, 1983), Chafe (1976, 1987), Ariel (1990, 2001) and Gundel et al. (1993). Ariel formalises the intuition that referring expressions mark varying degrees of mental accessibility of referents, as an inverse relation between linguistic encoding and accessibility. A new referent is inaccessible (i.e. inactive), and thus requires the most linguistic encoding (e.g. full NP + modifier), whereas a previously mentioned referent is considered accessible (active), and hence needs less linguistic encoding (with zero anaphora being the lowest in the scale). The degree of accessibility forms a scale referred to as the "accessibility hierarchy", as represented in (3) (see also the Givenness Hierarchy of Gundel et al. 1993).[3]

(3) Accessibility Hierarchy (Ariel 2001: 31):
     full name + modifier > full name > long definite description > short definite description > last name > first name > distal demonstrative + modifier > proximate demonstrative + modifier > distal demonstrative + NP > proximate demonstrative + NP > distal demonstrative > proximate demonstrative > stressed pronoun + gesture > stressed pronoun > unstressed pronoun > cliticized pronoun > verbal person inflections > zero

The accessibility/activation status not only influences the expression used for a referent, but also which information-structural function the referent is suited to play in the sentence. This is because the referent in the function of topic (which

---

[3]Apart from prototypical referents (things, people, places), the lexical content of verbs (events, states) may also be accessible to a higher or lower degree. Givenness of events may be expressed through ellipsis (e.g. 'Did you eat the last biscuit?'/'I did ~~eat the last biscuit~~'), and may influence the expression of the (often nominalised, Givón 2001) predicate as a topic. This is discussed further in literature on ellipsis (see e.g. Merchant 2001 on givenness of elided material), and predicate topicalisation and predicate doubling (see Güldemann & Fiedler 2022 for an overview of such constructions in Bantu). See also Hegarty et al. (2001) and Gundel et al. (2003) on discourse deixis and accessibility of entities introduced by a clause (e.g. 'We believe her, the court does not, and *that/it* resolves the matter', Gundel et al. 2003: 282).

will be covered in the next section) needs to be able to provide an anchoring point for the addressee to attach the new information to that is provided in a sentence, and therefore it cannot be completely new. This is captured in Lambrecht's (1994) Topic Acceptability Scale:

(4)   Topic Acceptability Scale (Lambrecht 1994: 165)
      active > accessible > unused > brand-new anchored > brand-new
      unanchored

      more acceptable   *less acceptable*

Two main information-structural functions have been proposed and widely accepted within the field of information structure (although see Matić & Wedgwood 2013 and Ozerov 2018, 2021 for critical views): *topic* and *focus*. While the function of a referent and its activation status interact, they are crucially different aspects. We discuss each function in more detail in the following sections.

3.2 Topic

In terms of information structure, most sentences can be split into a *topic* and a *comment*.[4] The comment, in turn, may contain a focused constituent, with the rest as background, as represented in Figure 2 (see also Neeleman et al. 2009, for example, for a separation in this way).

topic	comment	
	focus	background

Figure 2: Functions in the clause

The topic can be defined as "what the sentence is about" (Reinhart 1981), or the "spatial, temporal, or individual framework within which the main predication holds" (Chafe 1976: 50). This can be said to hold for all subtypes of topic. The topic-comment structure helps the addressee incorporate the new information into their knowledge, where the topic serves as an anchor to attach the comment to previously known information. Consider example (5): the question in (5a) targets the action of the 'woman', who represents the given information and functions as a topic. In the answer to this question in (5b), the topic referent 'woman' is expressed by a pronoun *ndé*, and the VP is the part that answers the content

---

[4]This concerns the sentence topic, and not the discourse topic. The latter refers to the broader topic of the paragraph or conversation; see Lambrecht (1994).

question, therefore forming the comment (and the focus in this case), expressed in SVO word order corresponding to Topic-Comment order.

(5) Teke-Kukuya (Chapter 3, Li 2025 [this volume])
Topic-comment structure (VP focus)

a. Mu-kái   kí-má   ká-sî?
1-woman 7-what 1SM.PST-do.PST
'What did the woman do?'

b. [Ndé]_{TOPIC} [á-búnum-i   baa-ntaba]_{COMMENT}.
1.PRO   1SM.PST-feed-PST 2-goat
'She fed the goats.'

Different subtypes of topic have been proposed. These subtypes are seen as conceptually distinct and this conceptual difference may correspond to differences in expression. At least in part, these different subtypes can be understood by the interaction of a referent's activation status and its topic function. When the referent is highly active and has already functioned as a topic before the current sentence, this is called a *familiarity topic* or *continuing topic* (e.g. Frascarelli & Hinterhölzl 2007, Givón 1983, Reinhart 1981). In all languages in our sample, such topics are preferably expressed through a pronoun (i.e. a reduced form). This is illustrated for Tunen in (6), where the class 1 subject marker *a* is used in the second and third sentences to refer to the personified referent 'chicken', which functions as a topic in each sentence and was previously introduced by a full noun phrase *miɔkɔ́* 'chicken'.

(6) Tunen (database Kerr)

u bótólókiə ɔ hité[!]yí hinó, miɔkɔ́ a ná hɛ́áná,
ɔ   bótólókiə ɔ   hité^Lyí   hɛ-nó **miɔkɔ́** a ná hɛ́áná
PREP begin   PREP DEM.DISC.EMPH.19 19-day **9.chicken** 1SM PST2 become
'À partir de ce jour, la poule est devenue -'
'From this day on, the chicken has become -'

a níŋə́kə na bɛ́ndɔ ɔmbél,
a   níŋɔ́-aka na bɛndɔ   ɔmbéla
1SM live-DUR with 2.person 3.house
'Elle vive avec les hommes à la maison.'
'She lives with humans in their homes.'

a bɛɔnɔ́ nɛakak,

a bɛ-ɔnɔ́ nɛaka-aka
1sm 8-egg make-DUR

'Elle pond des œufs ;'
'She lays eggs;'

bɛ́ndɔ bá nɛak.

bɛndɔ bá nɛa-aka
2.person 2sm eat-DUR

'les hommes (les) mangent.'
'people eat (them).'

Naturally, the topic must sometimes change. New topics like this are called *shift topics* (Erteschik-Shir 2007). A shift topic is illustrated above in (6) where in the last phrase the topic (and subject) shift from the chicken to *bɛndɔ* 'the people' and is expressed with a full noun phrase, i.e. using more linguistic material. We can see the same for Kîîtharaka in (7), where the topic (and subject) shifts from the referent 'the wife' to the referent 'the stick'. The stick therefore forms the shift topic, and because of its relatively less accessible status, *mûti* 'stick' is marked with a demonstrative, rather than using a more reduced encoding.

(7) (Previous story: Very long ago there was a man who had a wife and a daughter. They lived well until one day…) Kîîtharaka (database Kanampiu & Van der Wal)

Mwekûrû akiuna nkû aakûrûtwa i mûtî.

mû-ekûrû a-kî-un-a n-kû a-a-kûrût-w-a ni
1-wife 1sm-DEP-break-FV 9-firewood 1sm-PST-scratch-PASS-FV FOC
mû-tî
3-stick

'The wife was injured by a stick while collecting firewood.'

Mûti ûyû n'wari na cûmû.[5]

mû-tî ûyû ni-w-a-rî na cûmû
3-stick 3.DEM.PROX FOC-3sm-PST-be with poison

'This stick was poisonous.'

A third type of topic is a *contrastive topic* (Büring 2016, Repp 2010, Vermeulen 2012), which occurs when two topics are compared – even if one of them may be

---

[5]Note that the prefix *ni-* scopes over the verb and is independent of the initial phrase being the topic – see for further details Chapter 4 on Kîîtharaka, Kanampiu & van der Wal 2025 [this volume]

implicit. Asiimwe & van der Wal (2021) show for Rukiga that contrastive topics may be marked by a particle, as illustrated in example (8) with the class 11 particle *rwo*. This particle indicates that there is a contrast with other topical referents: here, when talking about various areas to be weeded (the topic referents), the particle indicates that the predicate holds for the banana plantation, but perhaps not for the maize garden or the sorghum field.

(8) Rukiga (Asiimwe & van der Wal 2021: 11)
     (Context: They have not weeded the maize garden, or there is another
     task they have not done)
     Orutookye **rwó** báárubágara.
     o-ru-tookye   **ru-o** ba-aa-ru-bagar-a
     ᴀᴜɢ-11-banana.plantation **11-ᴄᴍ** 2ꜱᴍ-ɴ.ᴘꜱᴛ-11ᴏᴍ-weed-ꜰᴠ
     'As for the banana plantation, they have weeded it.'

Contrastive topics seem to be connected to mirativity, that is, the expression of surprise (for the speaker or addressee), and also to intensive and depreciative readings. In the BaSIS languages, we see these interpretations for topic doubling (see Section 4.3), and for the use of the contrastive topic marker *-o*, as in (9). Recent research also shows a mirative interpretation of object-marker doubling, see Langa da Câmara et al. (2024), Lippard et al. (2023), Sikuku & Diercks (2021), Kerr & van der Wal (2023).

(9) Kîîtharaka (Chapter 4, Kanampiu & van der Wal 2025 [this volume])
     (Has it rained that much?)
     Mbúra **yoó** îkuúra.
     m-bura **î-o**   î-kû-ur-a
     9-rain   **9-ᴄᴍ** 9ꜱᴍ-ᴘʀꜱ-rain-ꜰᴠ
     'It has really rained.' (intensive interpretation)

In addition to these three types of topics, a sentence may be preceded by a *frame-setting* or *scene-setting* topic. Such topics are not always arguments of the sentence. For example, a temporal phrase sets the scene in (10), and example (11) shows a location, neither of which are arguments. Such topics may well be universal: Gundel (1988: 231) states that "Every language has syntactic topic constructions in which an expression which refers to the topic of the sentence is adjoined to the left of a full sentence comment." However, note that such "types" of topics only tell us something about the strategy to express them, i.e. the form, as their function as topics is still the same, namely restricting what the comment applies to.

(10)   Makhuwa (database Van der Wal)
       Mahíkw' éen' áalá tsinááthówa mivélo.
       ma-hiku ene ala   tsi-naa-thow-a   mi-velo
       6-day   INT 6.DEM.PROX   4SM-PRS.DJ-lack-FV 4-broom
       'These days there are not many brooms.'

(11)   Kirundi (database Nshemezimana & Van der Wal)
       Mu Rumonge amashuúre ntaruúgurura
       mu Rumonge a-ma-shuure nti-a-raa-ugurur-a
       18   Rumonge AUG-6-school NEG-6SM-INCP-open-FV
       'In Rumonge schools are not open yet.'

Topics can also combine and jointly restrict what the sentence is about. This is illustrated in (12), where the time frame is specified further with each topic phrase (marked by square brackets).

(12)   Kinyakyusa (Chapter 7, Lusekelo et al. 2025 [this volume])
       (Previous lines: Jackson has a dog which he loves very much. Also, he has
       a frog which he put in a bottle. It stayed and slept there.)
       [Akabalilo kamo] [pakilo] [Jakisoni n'mbwa jake] [bo bikulonda
       ukubuuka nkulambalala] bakikeetile ikyula.
       a-ka-balilo   ka-mo pa-kilo Jakisoni na   mbwa ji-ake   bo
       AUG-12-time 12-one 16-night 1.Jackson and 9.dog 9-POSS.1 when
       ba-ku-lond-a   u-ku-buuk-a   mu-ku-lambalal-a ba-ki-keet-ile
       2SM-PRS-want-FV AUG-15-go-FV 18-15-lie.down-FV 2SM-7OM-look-PFV
       i-ki-ula
       AUG-7-frog
       'One day, at night, when Jackson and his dog wanted to go to sleep, they
       looked at the frog.'

While speakers greatly prefer to partition sentences to first indicate the topic and then add a comment on that topic (a so-called *categorical* sentence), some sentences have no topic expression. These are *thetic* sentences, which present the information as one piece. Thetics have been said to be about the "here and now" – the topic referent is then the "stage topic" (Erteschik-Shir 1997, 2007, Gundel 1974). Thetics tend to occur when presenting a new entity, for example as in (13), or when announcing something out of the blue. In these cases, none of the referents are suitable to take on the role of topic. Since there is a well-known close association between subject and topic (Li & Thompson 1976), the (logical) subject in a thetic sentence is typically explicitly marked as *not* being the topic, for example by placing it in a postverbal position, as in (13), leading to departures

from the canonical word order. Theticity is discussed in detail in Sasse (1987, 1996, 2006).

(13) Kinyakyusa (database Lusekelo, Msovela, and Van der Wal)
(Previous lines: An old woman walked through the forest and was pierced by a thorn. She sat down and felt sorry for herself.)
Nakalinga akiindile **undumyana jumo**; ingamu jaake jo Lwitiko.
nakalinga a-kiind-ile   u-n-lumyana ju-mo i-ngamu   ji-ake
soon   1SM-pass-PFV AUG-1-young 1-one   AUG-9.name 9-POSS.SG
jo   Lwitiko
9.IDCOP 1.Lwitiko
'Soon there passed a certain boy; his name is Lwitiko.'

We thus see that the function of topic may be taken by referents that are at least identifiable, and that both topic and non-topic may be marked in languages, with the latter type visible in subjects of thetic sentences, which do not contain a topic-comment split.

3.3 Focus

In a categorical sentence (divided into topic and comment), the comment can be subdivided into the focus and the background, as seen above in Figure 2. We have informally referred to focus as the "new or contrastive" information, but a more formal definition is that focus "indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions" (Krifka & Musan 2012: 7; see Rooth 1992, 1996, 1985 for the Alternative Semantics approach). The constituent in focus thus triggers a set of relevant alternatives. For instance, in example (14), the referent 'bread' is in focus. This means that it triggers alternatives, i.e. other things that Asha could have baked, such as cake, chapati, doughnuts, and so on. The answer in (14) expresses that Asha baked bread as opposed to those alternatives – in other words, it selects bread as the relevant alternative out of the set of alternatives generated by the question word 'what?'.

(14) Kîîtharaka (database Kanampiu & Van der Wal)
(What has Asha baked?)
Áshá akáándíré [mûgááté]$_{FOC}$.
Asha   a-kaand-ire   mû-gaate
1.Asha 1SM-bake-PFV 3-bread
'Asha baked bread.'
Alternatives: {cake, bread, chapati, doughnuts, …}
Answer: {cake, **bread**, chapati, doughnuts, …}

There are two axes of variation in focus: the *scope* of focus and the *type* of focus. In terms of scope, there is a distinction between term focus (meaning focus on an argument or adjunct) and predicate-centred focus (PCF; following Güldemann 2003, 2009).[6] Within term focus, there can be focus on the whole noun phrase (whole NP focus) or focus on only a sub-element, such as a nominal modifier (sub-NP focus). Within PCF, there can be focus on the lexical value of the verb itself (the *state of affairs*), the truth value (also called *polarity* or *verum*), tense/aspect/mood (also referred to in terms of *operator focus*), or the whole verb phrase (*VP focus*). This is schematically represented in Figure 3 and illustrated in the examples below.
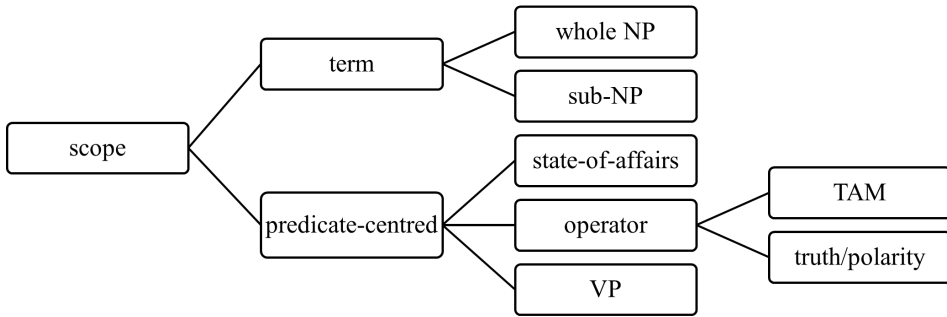
Figure 3: Different parts of the sentence which may be in focus (van der Wal 2021, building on Dik 1997, Güldemann 2009, Zimmermann 2016, Güldemann & Fiedler 2022: 540).

(15)　Term focus argument: Rukiga (Chapter 6, Asiimwe & van der Wal 2025 [this volume])

　　a.　Hélen atwire **ki**?
　　　　Helen　a-twar-ire　　ki
　　　　1.Hellen 1sm-take-pfv what
　　　　'What did Hellen take?'

　　b.　Hélen atwir' **ékikópo**.
　　　　Helen　a-twar-ire　　[e-ki-kopo]_{FOC}
　　　　1.Hellen 1sm-take-pfv　aug-7-cup
　　　　'Hellen took [a cup]_{FOC}.'

---

[6]We do not take Lambrecht's (1994) *sentence focus* to belong to this overview, as it concerns thetic sentences, discussed above.

(16) Term focus adjunct: Teke-Kukuya (Chapter 3, Li 2025 [this volume])

    a. Mwáana **munkí** ká-dzí   ntsúi?
       1.child   when 1SM.PST-eat.PST 1.fish
       'When did the child eat the fish?'

    b. Ndé ntsúi [**mu**   **ngwaalí**]_{FOC} ká-dzí.
       1.PRO 1.fish 18.LOC 9.morning   1SM.PST-eat.PST
       'S/He ate the fish [in the morning]_{FOC}.'

(17) Sub-NP focus: Rukiga (Asiimwe et al. 2023: 1297)
     (Which car should we take?)

    a.   Tu-twar-e   é-mótoka **é-y-ângye**.
       1PL.SM-take-SBJV AUG-9.car AUG-9-POSS.1SG
       'We take *my* car.'

    b. [#] Tu-twar-e   é-mótoka **y-angye**.
       1PL.SM-take-SBJV AUG-9.car 9-POSS.1SG
       'We take my car.'

(18) State-of-Affairs focus: Kîîtharaka (database Kanampiu & Van der Wal)
     I kûrííngá tû́ rííngiré ng'óombé – tûtíracíthaika.
     ni   kû-riinga tû-riing-ire   ng'-oombe tû-ti-ra-ci-thaik-a
     FOC 15-hit   1PL.SM-hit-PFV 10-cows   1PL.SM-NEG-YPST-10OM-tie-FV
     'We *hit* the cows, we didn't tie them.'

(19) VP focus: Changana (database Simango & Van der Wal)
     (Context: Picture of a man with a pumpkin in a wheelbarrow. Is this man playing football?)
     Ángábélí bó:lú, óxíxa kwe:mbe.
     a-nga-bel-i   bolu a-o-xix-a   kwembe
     1SM-NEG-play-NEG 5.ball 1SM-EXCL-unload-FV 5.pumpkin
     'No, he is not playing football, he is unloading a/the pumpkin.'

(20) Verum (Cicopi; Kerr & van der Wal 2023: 485)
     (You are not eating the cake that I bought. It'll go bad.)
     Kudya hâ:dyá.
     ku-dya hi-a-dy-a
     15-eat   1PL.SM-DJ-eat-FV
     'We *are* eating (it).'

Turning now to the type of focus, we need to distinguish between the semantic and pragmatic aspects of interpretation, following Krifka (2008), Zimmermann & Onea (2011). The semantic aspects affect what Krifka (2008: 249) calls the "common ground content" and therefore the truth-conditions of the utterance, whereas the pragmatic aspects concern the "common ground management", which does not affect the truth-conditions. It must be noted that it is not always clear whether an aspect of interpretation is rooted in the semantics or pragmatics. The types of focus are shown in the overview in Figure 4.
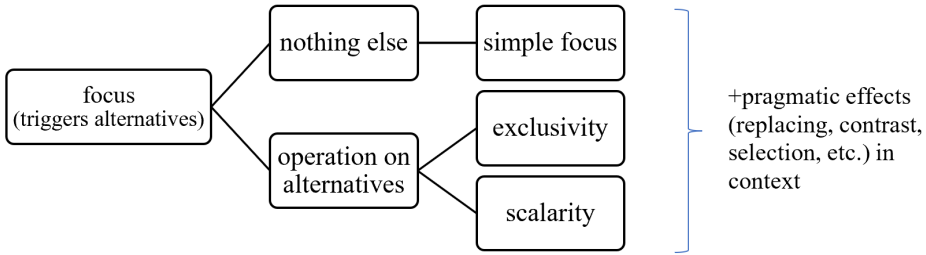
Figure 4: Types of semantic and pragmatic focus

The semantic aspect of focus concerns the set of alternatives. Under the Alternative Semantics approach to focus introduced at the start of this section, all focus types by definition trigger the generation of a set of alternatives. If nothing else happens, this is called *simple focus*, also known as "(new) information focus" (É. Kiss 1998) or "assertive focus" (Hyman & Watters 1984). For simple focus, the proposition is true for the focused referent, and there are alternatives for which we do not know if the proposition is true. This was illustrated using a question-answer context in (14) above, and is illustrated again for a simple focus context 'Who did you meet?' – 'I met Alex' in (21) below.

(21)   Alternatives: {Alex, Robin, Sam, Sacha, …}
        Simple focus: {**Alex**, Robin, Sam, Sacha, …}

On the other hand, an additional operation can be performed on the set of alternatives. This can be a scalar ordering of the referents, for example for degree of expectedness. Another option is that there are alternatives for which the proposition is false – in other words, not only is an alternative selected, but other alternatives are excluded. This type of focus is therefore called exclusive focus. This means that "there is at least one individual other than [the intended referent] for whom the proposition does not hold" (Kenesei 2006: 147). For example, in the sentence 'it is Alex that I met', the focus does not stop at identifying the

referent to which the proposition holds (Alex); it also excludes (at least some of) the possible alternatives (Robin, Sam, Sacha), implying that the proposition is false for them, as indicated by the strikethroughs in (22).

(22) Alternatives: {Alex, Robin, Sam, Sacha, …}
       Exclusive focus: {**Alex**, ~~Robin~~, ~~Sam~~, Sacha, …}

If the interpretation of the structure is that it excludes *all* the possible alternatives, it becomes *exhaustive* focus (which can be seen as a subtype of exclusive focus); see Szabolcsi (1981) for early discussion of exhaustivity; É. Kiss (1998) and many after her for discussion of the exhaustivity of Hungarian preverbal focus (e.g. Onea & Beaver 2009); and see Horn (1981), Drenhaus et al. (2011), among many others (e.g. the work of Destruel, as in Destruel et al. 2015, Destruel & DeVeaugh-Geiss 2018) for exhaustivity in clefts.

(23) Alternatives: {Alex, Robin, Sam, Sacha, …}
       Exhaustive focus: {**Alex**, ~~Robin~~, ~~Sam~~, ~~Sacha~~, ~~…~~}

Another aspect of focus is that a clause may contain a presupposition that there exists an entity for which the proposition is true, and this entity is then identified by the focus constituent. For instance, in the sentence 'the person I met is Alex', the phrase *the person I met* presupposes that there is a person whom the speaker met. This person is then identified as Alex through the focused NP 'Alex'. The role of focus in this structure can thus be said to be identificational. Note here that this is a different use of "identificational" than used by É. Kiss (1998), who uses this term for exhaustive focus.

On a pragmatic level, many more types of focus can be described, all dependent on the discourse context. We mention three types here (see Cruschina 2021, Dik 1997, Krifka 2008 among others for further types, e.g. confirmative, mirative, expanding). A first proposed type is *replacing focus*, also known as *corrective focus*. When the speaker suspects or knows that the addressee has a different referent in mind than the one intended by the speaker, they may try to correct this and replace the wrong referent with the intended one, thereby correcting the information.

A second "flavour" of focus is *contrastive focus*. This is possibly the most confusing term, since it has been used in two ways in the literature: firstly to express a contrast with a referent that is present in the context, and secondly to express contrast with referents in the set of alternatives. To some extent these overlap, of course, since referents in the context are also alternatives to the focused referent.

However, the two definitions of contrastive focus are not identical. As the contrast with the set of alternatives can be captured as the exclusion of those alternatives, we refer to that as exclusive focus (see also Byram Washburn 2013), and take contrastive focus to involve a pragmatic contrast with a contextual referent. What others have named "contrastive focus" can therefore be seen as *exclusive focus* in a *contrastive context*. For example in 'The children ate only *ugali* and the parents ate *rice*', the object ugali excludes alternative foods, and it contrasts in this context with rice. Note that, similarly, corrective/replacing focus is *exclusive focus* in a *corrective context*, as the referent that was replaced/corrected is excluded, e.g. the rice in 'The children ate ugali, not rice'. See Repp (2010), Molnár (2002) among others for further discussion on the notion of contrast, and Cruschina (2021) for a more detailed discussion on degrees of contrast. In our book, we use the term contrastive only in a pragmatic sense.

A third pragmatic type of focus is *selective focus*. This is found in a context where a set of alternatives is mentioned and one (or more) referents are selected from that set. Alternative questions such as 'Do you want coffee or tea?' establish a set of two referents (namely, coffee and tea) from which the addressee is to choose one. This choice, however, does not necessarily exclude the other referent and is therefore the same semantic type of focus as simple focus – the perceived exclusion of the non-selected alternative is thus in the pragmatics.

To sum up, focus relates to the generation of alternatives and, together with the background, forms the comment of a categorical sentence. Focus can be categorised in terms of the scope of the focus, with a basic split between term focus and predicate-centred focus, and in terms of the semantic and pragmatic type of focus. Semantically, all focus types trigger a set of alternatives. These alternatives may simply be there with no further operation upon them (*simple focus*), or they may be ordered (*scalar*) or (partly) excluded (*exclusive/exhaustive focus*). The pragmatic context may induce further variation in the interpretation, such as corrective focus, contrastive focus, and selective focus. Languages may express these different types of focus by different means.

4 Common constructions for information structure expression in Bantu

While the general aim of this chapter is to introduce the conceptual basics of information structure, it is also useful to introduce some constructions that occur in the Bantu languages covered in this book. We therefore turn now to zoom in on common constructions used for information structure expression in the Bantu

languages. In particular, we discuss cleft constructions, word order variation, and predicate doubling here. Clefts in particular are present in each of the BaSIS languages and always play a role in the expression of focus (see Gundel 1988 and Creissels 2020 for clefts as a universally available construction). There are three types of cleft constructions: *basic clefts*, *pseudoclefts*, and *reverse pseudoclefts*.

4.1 Cleft constructions

Basic cleft constructions[7] are bi-clausal structures that comprise a matrix clause consisting of a predicative nominal (typically headed by a copula) and a relative clause that is co-indexed with the predicative argument in the main clause (see Jespersen 1937: 1949; Lambrecht 2001: 467). In Tunen for example, a basic cleft is formed for subjects by the copula *á* preceding the focused subject, followed by a reduced relative clause. This is illustrated in (24), where the high tone in the subject marker indicates relativisation (contrasting with the low-toned non-relative class 1 subject marker *a*).

(24)   Tunen (Chapter 2, Kerr 2025 [this volume])
       (Which politician died?)
       Á Píɛ́l á ná wə.
       á   Píɛ́lɛ   á   ná   wə́
       COP 1.Pierre 1SM.REL PST2 die
       'It's Pierre who died.'
       COP [NP]$_{FOC}$ [reduced relative]

The pseudocleft and the reverse pseudocleft (see Collins 1991, Higgins 2015) share with the basic cleft that they contain a relative clause, a copula, and a highlighted element (Collins 1991: 482) or clefted constituent (as Hedberg & Fadden 2007 call it). The difference is in the order of these elements.

In pseudoclefts, the free relative clause presents the topic of the sentence while the post-copular nominal presents the focus (Hedberg & Fadden 2007, É. Kiss 1998). The structure in (25) below illustrates these two parts. Note that the free relative is headed in English by a question word, but in other languages a free relative is simply marked in whichever way a relative can be marked – in the case of Kîîtharaka this has developed from the distal demonstrative, and in other languages in the BaSIS sample it also typically is or derives from a demonstrative.

---

[7]We use this as the more language-neutral term for what has been called for English the "it-cleft" (Lafkioui et al. 2016); note that this is not to say that there is anything simpler about this cleft in comparison to other cleft types.

(25) Kîîtharaka (Chapter 4, Kanampiu & van der Wal 2025 [this volume])
(What did Karîmi kick?)

[Kî-rá	Karîmi	á-rííng-iré	]	í	[mû-bíírá	].
7-RM	Karimi	1SM-kick-PFV		COP	3-ball	
[	free	relative	]	COP	[NP	]
[	cleft	clause	]	COP	[clefted constituent]	

'What Karîmi kicked is a ball.'

In a pseudocleft, the free relative appears at the initial position. In a reverse pseudocleft, the order is inverted so that the NP (the clefted constituent) appears at the initial position and the relative follows the copula. The relative clause in the (reverse) pseudocleft is in fact a free relative, which functions as an independent phrase (see e.g. Šimík 2018). Therefore, (reverse) pseudoclefts can simply be seen as copular clauses of which one constituent is a free relative clause (Den Dikken 2006). An example of a reverse pseudocleft is given in (26) below.

(26) Kinyakyusa (Chapter 7, Lusekelo et al. 2025 [this volume])
(I did not find him/her in the river.)

[Ubibi jumo] jo [uju ambele].

u-bibi	ju-mo	jo	uju	a-m-pa-ile
AUG-1.grandmother	1-one	1.IDCOP	1.DEM.PROX	1SM-1SG.OM-give-PFV

'An old woman is the one who gave (him/her) to me.'

[NP]	COP	[free relative]
[clefted constituent]	COP	[cleft clause]

In a grammaticalisation process that has been shown to happen in languages all across the world (Harris & Campbell 1995, Heine & Traugott 1991), cleft constructions – particularly the basic cleft – develop from a biclausal construction into a monoclausal focus construction, as schematically represented in (27) (from van der Wal & Maniacky 2015: 9).

(27) [copula - NP] – [relative clause]   > [NP$_{FOC}$ Verb]
it is Maud – (the one) who made pancakes   > *Maud* made pancakes

Harris & Campbell (1995) define three stages in the development from biclausal to monoclausal structures, as shown in (28). The most obvious characteristics relevant for the languages in this book are the presence and form of the copula and the marking of relativisation, which are present in stage 1, but may disappear towards stage 3.

(28)   Development cleft > focus construction (Harris & Campbell 1995: 166)

> Stage 1:   The structure has all of the superficial characteristics of a biclausal structure and none of the characteristics of a monoclausal one.

> Stage 2:   The structure gradually acquires some characteristics of a monoclausal structure and retains some characteristics of a biclausal one.

> Stage 3:   The structure has all of the characteristics of a monoclausal structure and no characteristics of a biclausal one.

In various chapters, we will see properties indicating an in-between stage between biclausal clefts and monoclausal focus constructions, where the status of the erstwhile copula in a biclausal construction may later be analysed as a focus marker in a monoclausal construction (see discussion for Kîîtharaka and Kirundi, Chapters 4 and 5), or where the position has grammaticalised (as seen in Teke-Kukuya, Chapter 3).

On the interpretational side, the biclausal cleft typically expresses exhaustive focus, as a result of the combination of the relative clause and the predicative noun. The relative clause comes with a maximality presupposition: it refers to the complete set of referents for which the proposition holds. This maximal set of referents is then equated to the referent of the noun phrase. For example, in (29), we know that there is a unique individual who threw the oranges, and this individual is identified as Hare.

(29)   Makhuwa-Enahara (van der Wal 2009: 172)
       (Who has thrown oranges?)
       Namarokoló o-tthik-alé.
       1.hare.PRL   1-throw-PFV.REL
       'It was Hare who threw (them).'

In a monoclausal focus construction, the exhaustive focus interpretation that used to be the result of the whole biclausal construction is now associated with the focus marker (the erstwhile copula, for example) and/or a particular position. To illustrate, for Kîîtharaka, where the basic cleft has an exhaustive interpretation, Abels & Muriungi (2008) argue that the morpheme *ni* should no longer be analysed as a copula, but as a focus marker, shown in (30).

(30) Kîîtharaka (Chapter 4, Kanampiu & van der Wal 2025 [this volume])
Í mûgaáté Áshá agûkáánda.
ni   mû-gaate Asha   a-kû-kaand-a
FOC 3-bread   1.Asha 1SM-PRS-bake-FV
'It's bread that Asha is baking.'

While the aim of the chapters is not to analyse the underlying syntax in detail, but rather to provide a descriptive overview of the form-function mappings in each language, the diachronic reanalysis process is nevertheless useful to keep in mind. It is an interesting point of comparison between the Bantu languages, which vary in the extent of grammaticalisation into monoclausal focus constructions.

4.2 Word order

A second syntactic strategy used to express information structure is the manipulation of word order. Note here that we use the term "word order" in the chapters in this book, although in many instances we more specifically mean "constituent order": we are typically dealing with the order in which constituent *phrases* rather than words per se are expressed.

In almost all of the BaSIS languages (see Figure 1 above), word order plays a role in the expression of information structure, as we have shown in more detail in Kerr et al. (2023). Many eastern and southern Bantu languages show (some variant of) a topic-V-focus order (see e.g. van der Wal 2009, Yoneda 2011), resulting in a restriction on preverbal focus, using so-called "subject inversion constructions" to express theticity and/or subject focus, as illustrated in (31). The logical subject here appears in the "inverted" position after the verb because it is not the topic; the topic instead is the locative 'on the bridge' (as indicated in the context question), which therefore appears preverbally and determines subject agreement on the verb.

(31) Rukiga (Kerr et al. 2023: 2)
(What has happened on the bridge?)
A-ha   ru-tindo ha-a-raba=hó   e-motoka ny-îngi.
AUG-16 11-bridge 16SM-N.PST-pass=16 AUG-10.car 10-many
'On the bridge have passed many cars.'

If we think in terms of grammatical roles, these Bantu languages can be said to display SVO order with a (high) degree of flexibility (Bearth 2003, van der Wal 2015, Downing & Marten 2019, a.o.); for example, (31) shows a change to VS word order (i.e. locative inversion). However, if we describe word order in information-structural terms, as Good (2010) does, word order is not that flexible at all for some of the Bantu languages. The word order can in some languages be more insightfully described in terms of the following restrictions (i) non-topical/focal referents must be expressed after the verb, and (ii) topics are expressed preverbally, regardless of their syntactic or semantic role. In Kerr et al. (2023) we also show that three of the BaSIS languages have a dedicated focus position, defined as a position in which an element is always interpreted as focal: immediately before the verb (IBV) in Teke-Kukuya, immediately after the verb (IAV) in Makhuwa-Enahara, and clause-final in Kirundi (see example (1) above).

Nevertheless, there is a lot of variation in the use of word order to express information structure in Bantu, as can be seen in comparing the chapters of this volume (and see the discussion from a comparative perspective in Kerr et al. (2023)). One particular outlier is Tunen in the far northwest of the Bantu area, where subject inversion is impossible, and the same S-Aux-O-V word order is used to express a variety of information-structural interpretations. Tunen subjects, however, cannot be focused in their canonical position, and must be clefted, as we saw in (24) above.

4.3 Predicate doubling

A third type of construction that we found in the majority of languages is predicate doubling. These are constructions in which the predicate appears once in an inflected form and once in a non-finite form. There are three types of predicate doubling (see also Güldemann & Fiedler 2022, who use slightly different terminology): *cleft doubling*, *topic doubling*, and *in-situ doubling*. In cleft doubling, the non-finite form (typically an infinitive in class 15) forms the clefted constituent, as in (32), which is followed by the inflected (relative) form of the same predicate. This type of predicate doubling is typically used to express state-of-affairs focus.

(32) Kîîtharaka (database Kanampiu & Van der Wal)
      (Context: Someone thinks that you sang or cooked.)
      I kûrí̂ má tûrîmiré.
      ni   kû-rîma tû-rîm-ire
      FOC 15-dig   1PL.SM-dig-PFV
      'We (only) dug.'

Topic doubling involves an initial non-finite predicate, again typically an in-finitive, expressed as a contrastive topic, followed by the comment containing the same predicate. This can be used in verum contexts, as illustrated in (33), and in contexts where a contrast is perceived between the mentioned action and possible other, additional actions. This in turn may invite intensive and deprecia-tive interpretations, as described in this book for Kîîtharaka (Chapter 4), Kirundi (Chapter 5), Rukiga (Chapter 6), Kinyakyusa (Chapter 7), and Cicopi (Chapter 9). See also Jerro & van der Wal (in preparation) for a semantic-pragmatic analysis of topic doubling.

(33)   Cicopi (Chapter 9, Nhantumbo & van der Wal 2025 [this volume])
       (Context: A mother went out to work and when she returns, she can see
       that the children are weak. She asks the housemaid 'Are you cooking for
       these children?')
       Kubhí:ka hábhî:ka.
       ku-bhik-a   hi-a-bhik-a
       15-cook-FV 1PL.SM-DJ-cook-FV
       'We do cook.' (but they don't eat)

Finally, in-situ doubling is more variable both in form and in meaning. The non-finite form here follows the inflected form of the same predicate, but this can be an infinitive or a nominalised form. The latter is illustrated for Kirundi with the class 14 derivation *butunge* in (34). The interpretation for this construction in Kirundi indicates a prototypical, proper way of carrying out the action in the doubled verb. In other languages, in-situ doubling can function to express state-of-affairs focus or an intensive reading, for example.

(34)   Kirundi (database Nshemezimana & Van der Wal)
       (Does he keep the cows for someone else, or are they his?)
       Inka, arazitunze butunge.
       i-n-ka   a-zi-tung-ye   bu-tung-e
       AUG-10-cow 1SM-10OM-rear-PFV 14-rear-FV
       'Cows, he keeps them properly.'

There is, as expected, a lot of crosslinguistic variation in which of the three predicate doubling constructions are available (if any), in the non-finite form, and in the interpretation of the construction.

Cleft constructions, word order, and predicate doubling are only three of the morphosyntactic strategies in which information structure is expressed in the

Bantu languages. We provide more specific detail on these constructions in each chapter and hope that it will be helpful to have already seen their general structure in this section.

4.4 Interim summary

Summing up Sections 3 and 4, referents have a particular information status as more or less accessible, which influences their expression, and referents may further take one of two broadly-recognised information-structural functions: topic or focus. There are different types of topic and focus, which may require different linguistic strategies of expression. Common strategies used in the Bantu languages are cleft constructions, word order variation, and predicate doubling constructions, but there are many more. In the chapters in this book, we aim to determine which semantic and pragmatic interpretations the various linguistic strategies have in each language, working from form to function. As mentioned, we do this on the basis of a common methodology using diagnostics for information structure. The next sections present this BaSIS methodology (Section 5) and explain its key diagnostics used in this book (Section 6).
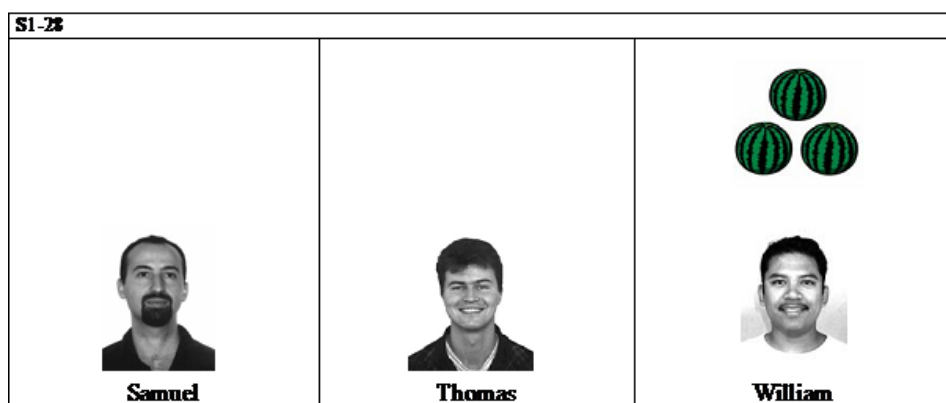
5 Methodology

For each language, apart from Kirundi where COVID-19 restrictions prevented fieldwork, we gathered data with at least three native speakers of the language, in locations where the language of interest was dominant in the community. Nevertheless, there is always the chance of influence from another language due to the use of a language of common communication for the elicitation process, as well as widespread multilingualism. The metalanguage used in the fieldwork was French for Tunen, Teke-Kukuya, and Kirundi; English for Kîîtharaka and Rukiga; Portuguese for Cicopi and Makhuwa-Enahara (as well as Changana); and Swahili and English for Kinyakyusa. Sessions for elicitation, as well as spontaneous discourse and narratives were recorded with audio (.wav files), and for Tunen also video (.mp4 files). The speakers all gave informed consent for the recordings and data gathering and received financial compensation for their time. Further information about the fieldwork is given in each individual chapter. The databases for all languages and Tunen recordings will soon be available in The Language Archive of the MPI (see the BaSIS collection[8]).

---

[8]https://hdl.handle.net/1839/2acf92c5-e5db-445b-bcdb-3b13b7e58f3f

The data gathering was conducted in a systematic way for each language, each researcher using the BaSIS methodology document that was developed over the course of the project. This methodology, further elaborated on in Section 6 below, draws on the Questionnaire of Information Structure – QUIS (Skopeteas et al. 2006), the Questionnaire on Focus Semantics (Renans et al. 2010), and the additional tests gathered in van der Wal (2016). The methodology document, which is freely available via the project website[9] and the Leiden Repository[10] (van der Wal 2021), grew during the project in three main ways. First, Part I was added, providing a basic introduction to understand the key concepts of information structure such that the methodology can be used by researchers new to information structure. Second, the rather artificial picture stimuli developed by the QUIS were gradually replaced by more natural pictures – compare Figure 5 and Figure 6. This change was made as the unnaturalness of the initial stimuli sometimes affected the fieldwork, for example with speakers indicating for Figure 5 that 'there are three melons above William's head'.

Figure 5: QUIS stimulus for corrective focus on numeral

The third addition to the methodology was Part III, where tests were added for Vergnaud licensing (abstract Case), as proposed by Sheehan & van der Wal (2018). This allows for investigation of nominal licensing, another goal of the BaSIS project (although not the focus of the chapters in this book).

---

[9]https://bantusyntaxinformationstructure.com/methodology/
[10]https://scholarlypublications.universiteitleiden.nl/handle/1887/3608096

Figure 6: BaSIS picture, which can be used as stimulus for corrective focus on numeral (e.g. "Does the woman have four knives?")

The methodology thus consisted of three parts: Part I providing a background on information structure so as to be consistent in conceptualisation and terminology, Part II providing an overview of all the diagnostics to elicit and test different notions of information structure (see Section 6), and Part III providing tests for nominal licensing. The data collected include both (semi-)spontaneous speech and elicitation data. The (semi-)spontaneous speech data took the form of recounted recipes, folk tales, the Frog Story, personal histories, free dialogues, and the QUIS map task. The elicitation data were collected for the purposes of systematic crosslinguistic comparison (as in the comparative work on word order by Kerr et al. 2023, and on truth expression by Kerr & van der Wal 2023). Additionally, elicitation tests are useful in providing negative data, which show which constructions and interpretations are *not* possible. Here, the BaSIS elicitation tests provide explicit information-structural context, allowing importantly for both the study of grammaticality (which constructions are well-formed in the language, with the symbol * in the data line indicating ungrammaticality) and felicity (which constructions are appropriate within the given discourse context, with the symbol # indicating infelicity).

The diagnostics in the methodology work in two directions: *function to form*, and *form to function.* An example of working from function to form is the commonly used question-answer test (also known as *question-answer congruence*), which investigates the expression of focus. As content question words are taken to be inherently focused, the answer to such a question forms the focus as well. Question-answer pairs can therefore be used to investigate focus expression. This test is illustrated for Cicopi in (35). Here, the translation of the question from Portuguese in (35a) shows us a strategy that is used to express an interrogative, and the answer to that question in (35b) also shows a focus strategy. We learn here that focus (function) on the theme argument can be expressed with the theme argument in a postverbal position (form) in Cicopi, both for interrogatives and for their answers.

(35) Cicopi (database Nhantumbo & van der Wal)

    a. Utóselá **cà:nì**?
        u-to-sel-a   **cani**
        2SG.SM-TO-drink-FV what
        '[What]$_{FOC}$ did you drink?'

    b. Nitóselá **sérvhejhá**.
        ni-to-sel-a   **servhejha**
        1SG.SM-TO-drink-FV beer
        'I drank [beer]$_{FOC}$.'

Working the other way, from form to function, is useful when a strategy has been identified and the details of its interpretation are of interest. For example, during the fieldwork we knew that Kinyakyusa has a CV nominal prefix, and wanted to investigate the interpretation of this prefix. The form-to-function tests in the methodology allowed us to test for focus and exhaustivity. If the CV prefix were a marker of exhaustive focus, meaning that this is the only referent for which the proposition is true ('he drank only *the big one*'), then it should be impossible to follow up with a clause stating the truth for another referent too and cancelling the exhaustivity ('also the small one'). This prediction turned out to be correct (36), and so we can conclude that, according to this test, the CV prefix (form) here is exhaustively focused (function).

(36) Kinyakyusa (van der Wal & Lusekelo 2022: 335)
    Anwile **jinywaamu** [#]n' iinaandi
    a-nw-ile   **ji**-nywamu na i-nandi
    1SM-drink-PFV EXH-9.big   and AUG-9.small
    'He drank (only) the big one [#]and also the small one.'

All the data collected through these tests, together with the relevant metadata, were entered into Online Language Databases (OLDs) for each language, which were accessed via the graphical user interface Dative[11]. The major benefit of Dative is that multiple users can access the database simultaneously and add data. This is crucial in a collaborative project, and this aspect worked very well.[12] Access to any of the databases can be granted by contacting Jenneke van der Wal and/or consulting the archival deposit in The Language Archive. An example of the Dative database layout and data/metadata input is given in Figure 7 and Figure 8 respectively.

Having introduced the basic structure of the methodology, let us now turn to consider in more detail how the methodology is applied for the investigation of information-structural concepts as introduced in Section 3 above. These sections together are intended to form the background to understanding the discussion and diagnostics in each of the individual chapters.

6 Information structure diagnostics

In determining the form-function mappings in the area of information structure, there are three steps:

(37)   How to diagnose information structure
  1. Find out how a function is expressed;
  2. Test the precise function of each form;
  3. Test whether the function found is inherent to the form, or associated pragmatically.

The first step, function to form, can for example be a question-answer pair, as illustrated in (35) above: we know that content question words are in focus, and that the part that answers the content question is also in focus, therefore the strategy that is used in the question and the answer is a strategy used to express focus. Note that this does not necessarily mean that it is a "focus strategy", as at this point we do not know the exact interpretation, nor whether focus is a necessary aspect of the strategy (see Matić & Wedgwood 2013 for further nuance on the latter point, and Ozerov 2022 for a critical discussion of the role of questions in information structure, seeing them as the product of diverse discourse processes).

---

[11]https://www.dative.ca/

[12]Other good aspects are the fact that Dative is quite intuitive to work with, and allows for elaborated combined search options and easy tagging of examples for properties of interest (e.g. "focus_obj" for object focus). A major downside at the time of writing is that it is still an online database, thus requiring internet connection and power.

Dative Forms Corpora Resources Analysis Help Account   Makhuwa

Browsing **all forms**. Viewing forms 1,801 to 1,900 of 1,946
(page 19 of 20)

100 items per page   16 17 18 **19** 20

× Advanced Search Smart Search   Advanced Search over Forms   Browse All

··· transcription ▾   regex ▾

Search Count Save

(1,801) ólé,   warumpyááwe   khúshélelíyá,   aá
         /ole   wa-rup-al-aawe   khu-sheleliya   aaa
         1.dem.dist   16-sleep-pfv.rel-poss.1 narr-dawn   ah

         kinrówá   waashínúni   vá
         ki-n-row-a   wa-shinuni   va/
         1sg.sm-prs.cj-go-fv 16-bird   16.dem.prox

         'He, the next day, (he said) Ah, I'm going bird hunting.'

(1,802) aá   okushé   nlénsó   'nna
         /aa   o-kush-e   n-lenso   nna
         ah   2sg.sm-take-sbjv 5-cloth   5.dem.prox

         wiítthúkele   mú
         w-ii-tthuk-el-e   mu/
         2sg.sm-refl-tie-appl-sbjv 18.dem.prox

         'Okay, take this cloth and tie it on yourself.'

(1,803) ólé   ohiítthúkélá   nlénsó   'nne
         /ole   o-h-ii-tthuk-el-a   n-lenso   nne/
         1.dem.dist 1sm-prfv.dj-refl-tie-appl-fv 5-cloth 5.dem.dist

         'He tied that cloth on himself.'

(1,804) masi   úle   aanítsúwela   étthú
         /masi   ole   aa-ni-tsuwel-a   etthu
         but   1.dem.dist 1sm.pst-hab-know-fv 9.thing

         yaaphééla   omwîirá,   ookúshá   nlénsó
         e-aa-pheel-a   o-n-iira   o-o-kush-a   n-lenso
         9-ipfv-want-fv   15-1om-do   1sm-prfv.dj-take-fv 5-cloth

         'nne
         nne/
         5.dem.dist

         'But (as) he knew what she wanted to do to him, he took the cloth,'

Figure 7: Screenshot of Dative for the Makhuwa-Enahara Online Language Database, showing transcription of each utterance with interlinear glossing and translation.

Figure 8: Screenshot of Dative for one form in the Makhuwa-Enahara OLD, showing part of the fields for data and metadata.

Continuing the same example, the second step (from form to function) might involve testing whether the strategy that is used in question-answer pairs is used for simple focus, or also exhaustive focus. If it is found that the interpretation is one of exhaustive focus, then the third step is to check whether this aspect of exhaustivity is necessarily present, or can for example be cancelled (indicating that it would be in the pragmatics and not an inherent, semantic aspect of inter-pretation associated with this form).

Regardless of which step or diagnostic is applied, it is essential that examples come with a discourse context. This is automatically the case for spontaneous examples and longer stretches of discourse or narrative, but should also be pro-vided for elicited examples, so that the reader and researcher can identify how the form of the utterance relates to the information structure as visible through the discourse context. We therefore introduced a discourse context during the elicitation, either by verbal means or through use of picture stimuli, rather than working from isolated translation of the metalanguage (in which case it can be difficult to know exactly what discourse context the speaker has in mind).

In what follows, we explain and illustrate some of the diagnostics for topic (in Section 6.1), focus (in Section 6.2), and thetics (in Section 6.3). The full set of diagnostics and explanations can be found in the methodology document (van der Wal 2021); see also the QUIS (Skopeteas et al. 2006), as well as Renans et al. (2010), van der Wal (2016) and Aissen (2023).

6.1 Topic diagnostics

While there is a crosslinguistic tendency for topics to be expressed sentence-initially (Gundel 1988, Aissen 2023, among others), the identification of topics has received less attention than the identification of focus.

Reinhart (1981) makes various suggestions for topic tests, among which is the periphrasis test: If a sentence can be paraphrased by 'as for X' or 'I say about X that …', the referent that X refers to is indeed the topic. As a second test, Ameka (1991) shows that topic-marking strategies are incompatible with focus and should therefore not co-occur with inherently focal question words or (exhaustive) focus markers.

A third practical test that has turned out to be useful for identifying topics is testing for indefinites (Reinhart 1981: 66, referring to Firbas 1966). Because the topic is the point to which the information in the comment is anchored, this point must be at least identifiable, and therefore cannot be an indefinite non-specific referent. In other words, if an element is identified as a non-specific indefinite, it cannot also be a topic. Consider as an illustration examples (38) and (39). In Teke-Kukuya, topical elements tend to occur in the preverbal domain, and the preverbal NP 'person' in (38a) must be interpreted as definite, while the indefinite reading for the context is impossible. In (38b), when the word 'person' occurs postverbally, either a definite or indefinite reading is possible.[13]

(38) (In)definite test: Teke-Kukuya (Chapter 3, Li 2025 [this volume])
      (Context: Your mum and you are entering a dark hall, you are walking in front and your mum asks you from behind if you saw anyone in the hall.)

   a. # Me   mbuurú ka-á-mún-i   ni.
       1SG.PRO 1.person NEG-1SG.PST-see-PST NEG
       'I did not see the person/*anyone.'

   b. Me   ka-á-mún-i   mbuurú ni.
       1SG.PRO NEG-1SG.PST-see-PST 1.person NEG
       'I did not see the person/anyone.'

---

[13]Note that this is not the immediate before verb focus position, as can be seen in the position of negation (see Chapter 3, Li 2025 [this volume], and Li 2024).

In contrast, the preverbal subject 'person' in Kinyakyusa (39) can be interpreted as indefinite and non-specific, thereby showing that preverbal constituents do not necessarily refer to topics.

(39) (In)definite test: Kinyakyusa (Chapter 7, Lusekelo et al. 2025 [this volume])
Linga siku **umundu** linga ikukubuula gwinogonengepo.
linga siku u-mu-ndu linga i-ku-ku-buul-a
COND 9.say AUG-1-person COND 1SM-PRS-2SG.OM-tell-FV
gw-inogon-ang-e=po
2SG.SM-think-?-SBJV=16
'If during another day someone tells you something, you must think.'

Weakly quantified NPs such as 'few' in (40), equally resist topicalisation, as they too behave as indefinites. The example in (40) shows that Makhuwa also has a subject position preverbally which can be filled by non-topics.

(40) Makhuwa-Enahara (van der Wal 2009: 176)
Epaáwú vakhaání yoo-khúúr-íy-a.
9.bread few 9SM.PFV.DJ-chew-PASS-FV
'Little bread was eaten.'

A final way to test topics is to distinguish subjects from topics through considering not only categorical sentences (with topic-comment) split but also thetic sentences (which do not have a topic). Investigating thetics can thus help to show whether preverbal noun phrases are best described in terms of topicality or subjecthood. For example, the availability of initial NPs in Tunen thetics shows that the word order is best described as Subject-Aux-O-V, rather than Top-Aux-O-V (Chapter 2, Kerr 2025 [this volume]).

6.2 Focus diagnostics

For the precise interpretation of focus, we first want to test whether a construction is about focus at all: if it does express focus, then we would expect it to allow content question words, answers to those questions, and constituents that can associate with the focus-sensitive particle 'only'. Furthermore, we would expect "unfocusable" constituents to not be tolerated in this strategy. Such "unfocusables" are elements for which no alternative set can be generated: in our methodology, we use parts of idioms and cognate objects as tests for such unfocusable elements. The logic here is that there are no alternatives for parts of idioms in their idiomatic interpretation, as this interpretation is dependent on

other parts of the idiom. For example, while the English idiom 'she lost her marbles' can be interpreted as 'she has gone crazy', focusing the object in a cleft 'it is her marbles that she lost' only leaves the literal interpretation of someone losing actual glass marbles (as opposed to alternatives such as a handkerchief or pencils), no longer allowing the idiomatic interpretation (for more detail, see van der Wal 2021). The same logic holds for cognate objects, such as 'dance a dance' or 'dream a dream': there are no valid alternatives that could be triggered (what else would you dream if not a dream?) and therefore the cognate object cannot be in focus (#'It is a dream she dreamed'). To illustrate this, the conjoint verb form in Kirundi cannot be used with a cognate object, as shown in (41), compared with the perfectly acceptable disjoint verb form in (41). This incompatibility shows the exclusive reading of the element following the conjoint verb form. Adding a relative clause to the object, in this case modifying 'dreams' by the relative clause 'that are not pleasant', makes it possible again to generate alternatives, such as dreams that are pleasant. This enables the generation of alternatives, and therefore allows the use of the conjoint verb form.

(41) Kirundi (Chapter 5, Nshemezimana & van der Wal 2025 [this volume])

DJ Naaróose indóoto.
    N-a-a-róot-ye   i-n-róoto
    1SG.SM-N.PST-DJ-dream-PFV AUG-9-dream
    'I dreamt a dream.'

CJ Naroose indoto #(zitari nziiza).
    N-a-róot-ye   i-n-róoto   zi-ta-ri   n-ziiza
    1SG.SM-N.PST.CJ-dream-PFV AUG-10-dream 10SM-NEG-be.REL 10-good
    'I dreamt dreams #(that were not pleasant).'

In a second step into the investigation of focus interpretation, we can test for exclusivity and exhaustivity. If a strategy is used for exclusive focus, this means that some or all of the alternatives should be excluded. We can test this by using items that do not allow exclusion and checking their acceptability in a focus strategy. There are at least five ways to test exclusivity:

(42) Tests for exclusive and exhaustive focus
      1. alternative questions;
      2. universal quantifiers 'all' and 'every';
      3. the focus-sensitive scalar particle 'even' and additive 'also';
      4. indefinite noun-phrases;
      5. a 'mention some' context.

We discuss these briefly here. First, when responding to an alternative question, i.e. a question that mentions two or more alternatives, an answer typically not only indicates for which referent the statement is true (e.g. Emily in (43)), but also that alternatives are excluded (in this example, Hamida), thus differing from the answer to a simple content question in which only selection of an alternative is involved (simple focus).

(43) Rukiga (database Asiimwe & Van der Wal)
      Kéék' ogihíire Émíri nínga Hamída?
      keeki o-gi-h-ire   Emily   nainga Hamida
      9.cake 2SG.SM-9OM-give-PFV 1.Emily or   1.Hamida
      'Did you give Emily or Hamida a cake?'

      Naagihíire Émirí atárí Hamída.
      n-gi-h-ire   Emily a-ta-ri   Hamida
      1SG.SM-9OM-give-PFV 1.Emily 1SM-NEG-be 1.Hamida
      'I gave it to Emily not Hamida.'

The second test for exclusivity uses universal quantifiers 'all' and 'every'. These quantifiers are incompatible with exclusive focus, because all referents are included and therefore there can be no exclusion of alternatives from the set. A similar reasoning holds for the additive particle 'also', which indicates that more instantiations of the action/state described in the predicate have occurred for different referents, therefore making the referent it modifies non-exhaustive. Similarly, the scalar additive particle 'even' presupposes that more instantiations of the action/state described in the predicate have occurred, and in addition expresses that the object modified by 'even' is the least likely in the set of contextually relevant alternatives to make the proposition true. Therefore, none of the alternatives are excluded, and a DP modified by 'even' is predicted to be infelicitous if a focus strategy is inherently exclusive. The unacceptability of the universal quantifier and the particle 'even' in a pseudocleft construction are shown in the examples in (44) and (45), respectively. This shows that the pseudocleft in Kinyakyusa has an exclusive interpretation.

(44) Kinyakyusa (Chapter 7, Lusekelo et al. 2025 [this volume])
      Ifi aagogile Kato fitana (*fyoosa).[14]
      ifi   a-a-gog-ile   Kato   fi-tana fi-osa
      8.DEM.PROX 1SM-PST-kill-PFV 1.Kato 8-cup   8-all
      'What Kato broke is (*all) cups.'

---

[14]Predication is expressed here by the absence of the augment – no segmental copula is needed.

(45) Uju nalyaganiile nagwe jo (*joope) n'uFrida.

uju   n-ali-aganil-ile   na-gwe   jo   j-oope na
1.DEM.PROX 1SG.SM-PST-meet-PFV with-1.PRO 1.IDCOP 1-even and
u-Frida
AUG-1.Frida
'The one I met is (*even) Frida.'

However, the incompatibility can be remedied by specifying a set of alternatives for the universally quantified DP, either by contrasting with another whole set, e.g. 'he broke all the cups, not all the plates', or by adding a restrictive relative clause, e.g. 'he broke all the cups *that are in the cupboard*, not the ones on the table'.

Indefinite non-specific phrases such as 'someone', 'anything', or 'nobody' cannot exclude alternatives: any (or no) referent will satisfy. To illustrate, when saying 'I need to eat something', the natural non-specific interpretation is that the speaker is hungry and does not care what they eat: anything will do out of the set of edible alternatives. All alternatives are therefore included. If a strategy expresses exclusive focus, it is thus predicted to be incompatible with the inclusive nature of the non-specific indefinite. This prediction is borne out for the conjoint and disjoint verb forms in Makhuwa-Enahara: the non-specific interpretation is possible after the disjoint verb form as in (46a), but unacceptable after the conjoint verb form as in (46b). The conjoint verb form is therefore said to express exclusive focus on the element directly following the verb (van der Wal 2011). In order to fit the noun *ntthu* 'person' into the exclusive interpretation, it interpreted as a type (a human being), as in (46c): other types can now be excluded, and the conjoint form is acceptable.

(46) Makhuwa-Enahara (van der Wal 2011: 1740)

a. DJ Koḿwéha ńtthu.
        ki-o-n-weh-a   n-tthu
        1SG.SM-PFV.DJ-1OM-look-FV 1-person
        'I saw someone.'

b. CJ #Ki-m-weh-alé   n-tthú.
        1SG.SM-1OM-look-PFV.CJ 1-person
        int.: 'I saw someone.'

c. CJ Ki-m-weh-alé   n-tthú, nki-weh-álé   enáma.
        1SG.SM-1OM-look-PFV.CJ 1-person NEG.1SG.SM-look-PFV 9.animal
        'I saw a person/human being, not an animal.'

A fourth test for exclusivity is the "mention some" context. Van der Wal (2016: 270) writes that "Instances of non-exhaustive focus are found in answers to so-called mention-some questions, where the context of the question does not require, or even allow for an explicit listing of all the true alternatives" (see also Abels & Muriungi 2008).

For example, if you can usually buy milk or tomatoes in various places, there is no single correct answer to a question 'Where can I buy milk?'. An exhaustive focus strategy is thus predicted to be infelicitous in the answer to a mention-some question. The test can therefore be used to discover which strategy is used in non-exhaustive focus (function to form), but can also be used to test suspected exhaustive focus strategies in non-exhaustive contexts (form to function).

There are a number of other tests in our methodology to test exclusivity and exhaustivity, for example the exact reading of numerals under exhaustive focus and the answers to incomplete or overcomplete yes/no questions. We illustrate here with Kinyakyusa numerals for the CV prefix and refer to the methodology document[15] for further tests and explanations. The meaning of numerals has been taken to have an underspecified interpretation either as the exact amount, or as a lower boundary 'at least this amount' (Horn 1972, Levinson 2000). However, in exhaustive focus, numerals lose their upward entailing quality and refer only to the exact quantity, because other amounts are excluded. In Kinyakyusa, a numeral in a DP with a CV marker is interpreted as the exact amount, as illustrated by the infelicity of the follow-up 'maybe more' in (47). The CV marker is concluded (also on the basis of other tests) to have an exhaustive interpretation (van der Wal & Lusekelo 2022 and Chapter 7).

(47) Kinyakyusa (van der Wal & Lusekelo 2022)
      Singuku ntandatu syo isi syalyusigwe (#pamo n' iisiingi).
      si-nguku   ntandatu syo   isi
      EXH-10.chicken 10.six   10.IDCOP 10.DEM.PROX
      si-ali-ul-is-igw-e   pamo na i-si-ngi
      10SM-PST-buy-CAUS-PASS-FV maybe and AUG-10-other
      'It's six chickens (exactly) that were sold (#maybe more).'

A final important focus diagnostic concerns the detection of a presupposition. Whether a presupposition of existence is inherent to a linguistic strategy can be tested by asking a content question using that strategy and establishing whether the question can felicitously be answered by 'nothing' or 'nobody', i.e. the empty

---

[15]https://scholarlypublications.universiteitleiden.nl/handle/1887/3608096

set. We are thus testing the strategy used in the question here. If the strategy presupposes that there is a referent for which the proposition is true, then logically the answer cannot deny that. Van der Wal & Namyalo (2016) give the following examples from Luganda, showing that the postverbal focus strategy does not have a presupposition, as the SVO object question in (48) can readily be answered by 'nobody', whereas the strategy formed by the question word in initial position followed by the marker *gwe* does have this presupposition, considering the infelicity of the answer in (49).

(48)   Luganda (van der Wal & Namyalo 2016: 360)

    Q: W-á-kúbyé   ání?
           2SG.SM-PST-hit.PFV who
           'Who did you hit?'

    A: Te-wá-lî.
           NEG-16SM-be
           'Nobody.', lit. 'There is not.'

(49)   Luganda (van der Wal & Namyalo 2016: 360)

    Q: Aní gw-e w-á-kúbyê?
           who 1-FOC 2SG.SM-PST-hit.PFV
           'Who is it that you hit?'

    A: [#] Te-wá-lî.
           NEG-16SM-be
           'Nobody.', lit. 'There is not.'

Each of these tests, as well as additional ones, will come back when necessary in each chapter to show the precise interpretations of the strategies for the expression of information structure.

6.3 Thetics

Testing the properties of thetic sentences (sentences without a topic-comment split) is more challenging than testing for topic and focus, as thetics quite often use a strategy that is also used for other interpretations. It is therefore difficult to directly test for their thetic interpretation. For example, the canonical word order can also be felicitous in a thetic context in some languages, and subject inversion constructions are often used for thetics, but may also be used for narrow focus on the subject. The latter example of overlap is illustrated in (50): default

agreement inversion (VS word order with class 17 default agreement) can be used in Changana both in a thetic context out of the blue, as in (50a), and also with narrow subject focus, as shown for the subject modified by exhaustive 'only' in (50b).

(50) Changana (van der Wal & Simango 2022: 236, 238)

    a. Kú-w-é   mú-ya:ki.
       17SM-fall-PFV.CJ 1-bricklayer
       'A bricklayer fell / There fell a bricklayer.'

    b. Kú-lúz-e   ntsé:ná kókwa:na.
       17SM-lose-PFV.CJ only   1.grandparent
       'Only grandpa passed away.'

We looked for thetics by using prototypical thetic contexts, asking speakers to describe a picture, describe the weather, or to imagine what would be said in an out of the blue (or *hot news*) context (e.g. "There is a village on a remote part of the river, where hardly any boats come to. One morning, a child comes running from the riverbank, shouting about the arrival of a ship. What does the child shout?"). It is often said that the answer to a question 'what happened?' is necessarily thetic, but this is not a watertight diagnostic for theticity. Because some referents are permanently or situationally available, these can be pushed into a topic function (Givón 1983, Erteschik-Shir 2007). For example, speech participants and "always available" referents such as 'the moon' or 'the newspaper' can always form the topic, even if they have not been mentioned in the previous discourse. Since speakers have a very strong preference to form categorical sentences containing a topic and comment, it is likely that such permanently available referents function as topics even in a context that might otherwise favour a thetic statement. That is, the answer to a question 'what happened?' (which does not present any referent as active) typically elicits a thetic answer, but if a referent is permanently available, it may be coded as a topic even in this environment.

6.4 Section summary

In this section we have only briefly touched upon the main diagnostics and their logic in application for determining the expression and precise interpretation of information structure in the BaSIS languages. Additional information is provided in each chapter, as needed in order to understand the diagnostics applied for each individual language. We refer the interested reader to the methodology

document[16] for more extensive and detailed information on the diagnostics as well as further examples, hoping that the brief overview here may help to interpret the information in the chapters of this volume.

7 Reading guide

In preparing this book, we have attempted to co-ordinate the structure of each chapter so as to allow for cross-linguistic comparison. At the same time, each researcher has their own research interests, their own set of tests covered in the fieldwork, and was guided by the different nature of each individual language, meaning that the chapters also differ in content and focus.

Each of the chapters starts out with some basic information about the language, such as the demography, classification, and geography. As mentioned, we do not provide general grammatical information about the language (e.g. the noun classes or tense system), but straightaway enter into presenting and discussing the main morphosyntactic strategies involved in the expression of information structure, providing references to other sources in which more detailed background information is available on each language.

We structure the presentation for each language from form to function: we take the linguistic (mostly morphosyntactic) expression and describe which function it has. This has some drawbacks, as it does not allow one to easily search "how does language X express contrastive topics", for example. Nevertheless, this presentation was chosen because strategies can be underspecified or multifunctional in what they express, meaning that providing an as-complete-as-possible description of the strategy will be beneficial (and also considering Matić & Wedgwood's (2013) warning that not every strategy that is used for focus is actually a focus strategy). The chapters are therefore organised with the goal of understanding the functions of each form in a given language, with cross-linguistic comparison a secondary aim.

All descriptions and generalisations presented are illustrated with example sentences, mostly from our own fieldwork and sometimes from existing sources. As in this introductory chapter, the examples that are extracted from the Online Language Databases via Dative generally have 4 lines: the expression in the object language, a morpheme break, a gloss, and a free translation. When the surface morphophonology is transparent enough, we may give only 3 lines by merging the first line and the morpheme break. Surface tone is marked on the first line for all the languages except Kinyakyusa, which does not have tone, and Kirundi, where the standard Kirundi orthography for tone marking is used.

---

[16]https://scholarlypublications.universiteitleiden.nl/handle/1887/3608096

Although the chapters are quite extensive, aiming for as complete a description as possible based on the available data, there will always be interesting phenomena that are not covered. We mention some of these limitations here. Firstly, as the project concentrated on the morphosyntactic expression of information structure (believed to be of key significance in Bantu languages), less attention is paid to phonology and prosody, only mentioning it where relevant and/or known. Secondly, we stick to the level of the sentence and do not discuss what happens on the level of the discourse. In other words, we do not go into discourse particles and the overall organisation of discourse or narratives (with the exception of some discussion of referent expression over discourse, as related to referent accessibility). Finally, as stated before, the aim of the book is descriptive, and so theoretical analyses and modelling have been avoided as far as possible – the intention is for the chapters to form a stepping stone from which such formal analyses may be developed.

Despite the limitations in topics and languages we could cover in this book, we hope that the chapters contribute to a better understanding both of the individual languages and of information structure in general. We also hope that this book and the methodology document may inspire others to describe information structure in other languages too, within and outside of Bantu. This will in turn enable us to better understand the diversity of the Bantu family and make more robust crosslinguistic comparisons regarding how information structure is expressed.

Acknowledgements

Abbreviations and symbols

Numbers refer to noun classes except when followed by SG/PL, in which case they refer to persons. Commas indicate a pause.

*	ungrammatical	*(X)	the presence of X is obligatory
#	infelicitous in the given context		and cannot grammatically be omitted

(*X)	the presence of X would make the sentence ungrammatical	INCP	inceptive
		INT	intensifier
(X)	the presence of X is optional	N.PST	near past
AUG	augment	OLD	Online Language Database
Aux	auxiliary	OM	object marker
CJ	conjoint	PREP	preposition
CM	contrastive marker	PRL	predicative lowering
COP	copula	PRO	pronoun
DEM	demonstrative	PST	past
DEP	dependent conjugation	PST2	Tunen second-degree past tense (hodiernal)
DISC	discourse		
DUR	durative/pluractional	REL	relative
DJ	disjoint	RM	relative marker
EMPH	emphatic	SM	subject marker
EXH	exhaustive	TO	Cicopi perfective tense
FV	final vowel	Top	topic
IDCOP	identificational copula	YPST	yesterday past

References

Abels, Klaus & Peter Muriungi. 2008. The focus marker in Kîîtharaka: Syntax and semantics. *Lingua* 118(5). 687–731. DOI: 10.1016/j.lingua.2007.09.003.

Aissen, Judith. 2023. Documenting topic and focus. In Peter Jenks & Lev Michael (eds.), *Key topics in language documentation and description* (Language Documentation & Conservation Special Publication 26), 11–57. Honolulu, HI: Department of Linguistics, UHM. https://nflrc.hawaii.edu/ldc/sp26/.

Ameka, Felix Kofi. 1991. *Ewe: Its grammatical constructions and illocutionary devices*. Canberra: Australian National University. (Doctoral dissertation).

Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.

Ariel, Mira. 2001. Accessibility theory: An overview. In Ted Sanders, Joost Schilperoord & Wilbert Spooren (eds.), *Text representation: Linguistic and psycholinguistic aspects*, 29–88. Amsterdam: John Benjamins.

Asiimwe, Allen, Maria Kouneli & Jenneke van der Wal. 2023. Determiner spreading in Rukiga. *Linguistics* 61(5). 1285–1339. DOI: 10.1515/ling-2021-0027.

Asiimwe, Allen & Jenneke van der Wal. 2021. The multifunctionality of -o in Rukiga: Pronoun, contrastive topic, and mirative marker. *Nordic Journal of African Studies* 30(1). 26–26. DOI: 10.53228/njas.v30i1.758.

Asiimwe, Allen & Jenneke van der Wal. 2025. The expression of information structure in Rukiga. In Jenneke van der Wal (ed.), *The expression of information structure in Bantu*, 335–405. Berlin: Language Science Press. DOI: 10.5281/zenodo.14833614.

Bearth, Thomas. 2003. Syntax. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 121–142. London, New York: Routledge.

Büring, Daniel. 2016. (Contrastive) Topic. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure*, 64–85. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199642670.013.002.

Byram Washburn, Mary. 2013. *Narrowing the focus: Experimental studies in exhaustivity and contrast.* Los Angeles: University of Southern California. (Doctoral dissertation.)

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Charles N. Li & Sandra A. Thompson (eds.), *Subject and topic*, 27–55. New York: Academic Press.

Chafe, Wallace L. 1987. Cognitive constraints on information flow. In Russell Tomlin (ed.), *Coherence and grounding in discourse*, 21–51. Amsterdam: John Benjamins.

Collins, Peter Craig. 1991. Pseudocleft and cleft constructions: A thematic and informational interpretation. *Linguistics* 29(3). 481–520. DOI: 10.1515/ling.1991.29.3.481.

Creissels, Denis. 2020. Remarks on the grammaticalization of identificational clefts. *Linguistic Discovery* 17(1). 7–20.

Cruschina, Silvio. 2021. The greater the contrast, the greater the potential: On the effects of focus in syntax. *Glossa: A journal of general linguistics* 6(1). 1–30. DOI: 10.5334/gjgl.1100.

Den Dikken, Marcel. 2006. Specificational copular sentences and pseudoclefts. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, 292–409. Hoboken: John Wiley & Sons, Ltd. DOI: 10.1002/9780470996591.ch61.

Destruel, Emilie & Joseph DeVeaugh-Geiss. 2018. On the interpretation and processing of exhaustivity: Evidence of variation in English and French clefts. *Journal of Pragmatics* 138. 1–16. DOI: 10.1016/j.pragma.2018.09.009.

Destruel, Emilie, Daniel Velleman, Edgar Onea, Dylan Bumford, Jingyang Xue & David Beaver. 2015. A cross-linguistic study of the non-at-issueness of exhaustive inferences. In Florian Schwarz (ed.), *Experimental perspectives on presuppositions*, 135–156. Berlin: Springer.

Dik, Simon C. 1997. *The theory of Functional Grammar.* Berlin/New York: de Gruyter.

Downing, Laura J. & Larry M. Hyman. 2016. Information structure in Bantu. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure*, 790–813. Oxford: Oxford University Press.

Downing, Laura J. & Lutz Marten. 2019. Clausal morphosyntax and information structure. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 2nd edition (Routledge Language Family Series), 270–307. London & New York: Routledge.

Drenhaus, Heiner, Malte Zimmermann & Shravan Vasishth. 2011. Exhaustiveness effects in clefts are not truth-functional. *Journal of Neurolinguistics* 24(3). 320–337.

É. Kiss, Katalin. 1998. Identificational focus versus information focus. *Language* 74(2). 245–273.

Erteschik-Shir, Nomi. 1997. *The dynamics of focus structure*. Cambridge: Cambridge University Press.

Erteschik-Shir, Nomi. 2007. *Information structure: The syntax-discourse interface*. Oxford: Oxford University Press.

Firbas, Jan. 1966. Non thematic subjects in contemporary English. *Travaux Linguistiques de Prague* 2(1). 239–256.

Frascarelli, Mara & Roland Hinterhölzl. 2007. Types of topics in German and Italian. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure, meaning and form*, 87–116. Amsterdam: John Benjamins.

Givón, Talmy. 1976. Topic, pronoun and grammatical agreement. In Charles Li (ed.), *Subject and topic*, 149–188. New York: Academic Press.

Givón, Talmy (ed.). 1983. *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam: John Benjamins.

Givón, Talmy. 2001. *Syntax: An introduction*, vol. 2. Amsterdam: John Benjamins.

Good, Jeff. 2010. Topic and focus fields in Naki. In Ines Fiedler & Anne Schwarz (eds.), *The expression of information structure: A documentation of its diversity across Africa*, 35–68. Amsterdam: John Benjamins.

Güldemann, Tom. 2003. Present progressive vis-à-vis predication focus in Bantu: A verbal category between semantics and pragmatics. *Studies in Language* 27(2). 323–360. DOI: 10.1075/sl.27.2.05gul.

Güldemann, Tom & Ines Fiedler. 2022. Predicate partition for predicate-centred focus and Meeussen's Proto-Bantu "advance verb construction". In Koen Bostoen, Gilles-Maurice Schryver, Rozenn Guérois & Sara Pacchiarotti (eds.), *On reconstructing Proto-Bantu grammar*, 537–580. Berlin: Language Science Press.

Güldemann, Tom. 2009. Predicate-centered focus types: A sample-based typological study in African languages. Application for project B7 in the Collaborative Research Center (Sonderforschungsbereich/SFB 632) "Information Structure". Potsdam University, Humboldt University of Berlin, Free University of Berlin.

Gundel, Jeanette K. 1974. *The role of Topic and Comment in linguistic theory*. Austin, TX: University of Texas at Austin. (Doctoral dissertation).

Gundel, Jeanette K. 1988. Universals of topic-comment structure. In Michael Hammond, Edith A. Moravcsik & Jessica Wirth (eds.), *Studies in syntactic typology*, 209–242. Amsterdam: John Benjamins. DOI: 10.1075/tsl.17.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69. 274–307.

Gundel, Jeanette K., Michael Hegarty & Kaja Borthen. 2003. Cognitive status, information structure, and pronominal reference to clausally introduced entities. *Journal of Logic, Language and Information* 12(3). 281–299. DOI: 10.1023/A: 1024102420659.

Guthrie, Malcolm. 1948. *The classification of the Bantu languages*. Oxford: Oxford University Press for The International Africa Institute.

Hammarström, Harald. 2019. An inventory of Bantu languages. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 2nd edn., 17–78. London: Routledge.

Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2022. *Glottolog 4.7*. Leipzig: Max Planck Institute for Evolutionary Anthropology. DOI: 10.5281/zenodo.7398962. https://glottolog.org/.

Harris, Alice C. & Lyle Campbell. 1995. *Historical syntax in cross linguistic perspective*. Cambridge: Cambridge University Press.

Hedberg, Nancy & Lorna Fadden. 2007. The information structure of it-clefts, Wh-clefts and reverse Wh-clefts in English. In Nancy Hedberg & Ron Zacharski (eds.), *The grammar–pragmatics interface: Essays in honor of Jeanette K. Gundel* (Pragmatics & Beyond New Series 155), 49–76. Amsterdam: John Benjamins. DOI: 10.1075/pbns.155.05hed.

Hegarty, Michael, Jeanette K. Gundel & Kaja Borthen. 2001. Information structure and the accessibility of clausally introduced referents. *ZAS Papers in Linguistics* 27(23). 163–186. DOI: 10.1515/thli.2001.27.2-3.163.

Heine, Bernd & Elizabeth C. Traugott. 1991. *Approaches to grammaticalization*, vol. 1: Theoretical and Methodological Issues (Typological Studies in Language 19). Amsterdam: John Benjamins.

Higgins, Francis Roger. 2015. *The pseudo-cleft construction in English*. Cambridge MA: Routledge.

Horn, Laurence R. 1972. *On the semantic properties of logical operators in English.* University of California, Los Angeles. (Doctoral dissertation).

Horn, Laurence R. 1981. Exhaustiveness and the semantics of clefts. In Virginia Burke & James Pustejovsky (eds.), *Proceedings of NELS 11*, 124–142. Amherst, MA: GLSA.

Hyman, Larry M. & John R. Watters. 1984. Auxiliary focus. *Studies in African Linguistics, Supplement* 15(3). 233–273. DOI: 10.32473/sal.v15i3.107511.

Jerro, Kyle & Jenneke van der Wal. In preparation. *A contrastive topic analysis of Bantu predicate doubling.*

Jespersen, Otto. 1937. *Analytic syntax.* Copenhagen: Levin and Munksgaard.

Kanampiu, Patrick N. & Jenneke van der Wal. 2025. The expression of information structure in Kîîtharaka. In Jenneke van der Wal (ed.), *The expression of information structure in Bantu*, 199–269. Berlin: Language Science Press. DOI: 10.5281/zenodo.14833610.

Kenesei, István. 2006. Focus as identification. In Valéria Molnár & Susanne Winkler (eds.), *Focus as identification*, 137–168. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110922011.137.

Kerr, Elisabeth J. 2025. The expression of information structure in Tunen. In Jenneke van der Wal (ed.), *The expression of information structure in Bantu*, 53–122. Berlin: Language Science Press. DOI: 10.5281/zenodo.14833606.

Kerr, Elisabeth J., Allen Asiimwe, Patrick N. Kanampiu, Zhen Li, Ernest Nshemezimana & Jenneke van der Wal. 2023. Bantu word order between discourse and syntactic relations. *Linguistique et langues africaines* 9(1). DOI: 10.4000/lla.9496.

Kerr, Elisabeth J. & Jenneke van der Wal. 2023. Indirect truth marking via backgrounding: Evidence from Bantu. *Zeitschrift für Sprachwissenschaft* 42(3). 443–492. DOI: 10.1515/zfs-2023-2010.

Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3-4). 243–276.

Krifka, Manfred & Renate Musan. 2012. Information structure: Overview and linguistic issues. In Manfred Krifka & Renate Musan (eds.), *The expression of information structure*, 1–44. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110261608.1.

Lafkioui, Mena, Ernest Nshemezimana & Koen Bostoen. 2016. Cleft constructions and focus in Kirundi. *Africana Linguistica* 22. 71–106.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents* (Cambridge Studies in Linguistics 71). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511620607.

Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39(3). 463–516. DOI: 10.1515/ling.2001.021.

Langa da Câmara, Crisófia, Michael Diercks, Madelyn Colantes, Brendan Ly, Jackson Kuzmik & Hannah Lippard. 2024. An initial look at object marking in Cinyungwe. In James Essegbey, Brent Henderson, Fiona McLaughlin & Michael Diercks (eds.), *Pushing the boundaries: Selected papers from the 51–52 Annual Conference on African Linguistics*, 269–295. Language Science Press.

Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature.* Cambridge, MA: MIT Press. DOI: 10.7551/mitpress/5526.001.0001.

Li, Charles N. & Sandra A. Thompson. 1976. Subject and topic: A new typology of language. In Charles N. Li (ed.), *Subject and topic*, 457–490. New York: Academic Press.

Li, Zhen. 2020. A note on the functional passives in Teke-Kukuya (Bantu B77, Congo). In Jenneke van der Wal, Heleen Smits, Sara Petrollino, Victoria Nyst & Maarten Kossmann (eds.), *Essays on African languages and linguistics: In honour of Maarten Mous*, 267–286. Leiden: ASCL Occasional Publication. https://hdl.handle.net/1887/138531.

Li, Zhen. 2024. *Word order, agreement and information structure in Teke-Kukuya.* Leiden University. (Doctoral dissertation).

Li, Zhen. 2025. The expression of information structure in Teke-Kukuya. In Jenneke van der Wal (ed.), *The expression of information structure in Bantu*, 123–197. Berlin: Language Science Press. DOI: 10.5281/zenodo.14833608.

Lippard, Hannah, Justine Sikuku, Crisófia Langa da Câmara, Rose Letsholo, Madelyn Colantes, Kang (Franco) Liu & Michael Diercks. 2023. Emphatic interpretations of object marking in Bantu languages. *Studies in African Linguistics* 52(Supplement 14). 78–109.

Lusekelo, Amani, Jenneke van der Wal & Simon Msovela. 2025. The expression of information structure in Kinyakyusa. In Jenneke van der Wal (ed.), *The expression of information structure in Bantu*, 407–452. Berlin: Language Science Press. DOI: 10.5281/zenodo.14833616.

Maho, Jouni Filip. 2003. A classification of the Bantu languages: An update of Guthrie's referential system. In Derek Nurse & Gérard Philippson (eds.), *The Bantu Languages*, 639–651. London: Routledge.

Maho, Jouni Filip. 2009. *NUGL online: The online version of the new updated Guthrie list, a referential classification of the Bantu Languages.* http://goto.glocalnet.net/mahopapers/nuglonline.pdf.

Matić, Dejan & Daniel Wedgwood. 2013. The meanings of focus: The significance of an interpretation-based category in cross-linguistic analysis. *Journal of Linguistics* 49(1). 127–163. DOI: 10.1017/S0022226712000345.

Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis.* Oxford: Oxford University Press.

Molnár, Valéria. 2002. Contrast – from a contrastive perspective. In Hilde Hasselgård, Stig Johansson, Bergljot Behrens & Cathrine Fabricius-Hansen (eds.), *Information structure in a cross-linguistic perspective*, 147–161. Leiden: Brill. DOI: 10.1163/9789004334250_010.

Neeleman, Ad, Elena Titov, Hans Koot & Reiko Vermeulen. 2009. *A syntactic typology of topic, focus and contrast.* Berlin: De Gruyter Mouton. 15–52. DOI: 10.1515/9783110217124.15.

Nhantumbo, Nelsa & Jenneke van der Wal. 2025. The expression of information structure in Cicopi. In Jenneke van der Wal (ed.), *The expression of information structure in Bantu*, 511–566. Berlin: Language Science Press. DOI: 10.5281/zenodo.14833620.

Nshemezimana, Ernest. 2016. *Morphosyntaxe et structure informationnelle en kirundi: Focus et stratégies de focalisation.* Ghent: Ghent University. (Doctoral dissertation).

Nshemezimana, Ernest & Jenneke van der Wal. 2025. The expression of information structure in Kirundi. In Jenneke van der Wal (ed.), *The expression of information structure in Bantu*, 271–334. Berlin: Language Science Press. DOI: 10.5281/zenodo.14833612.

Onea, Edgar & David Beaver. 2009. Hungarian focus is not exhausted. *Semantics and Linguistic Theory* 19. 342–359. DOI: 10.3765/salt.v19i0.2524.

Ozerov, Pavel. 2018. Tracing the sources of information structure: Towards the study of interactional management of information. *Journal of Pragmatics* 138. 77–97. DOI: 10.1016/j.pragma.2018.08.017.

Ozerov, Pavel. 2021. Multifactorial Information Management (MIM): Summing up the emerging alternative to information structure. *Linguistics Vanguard* 7(1). DOI: 10.1515/lingvan-2020-0039.

Ozerov, Pavel. 2022. Alternatives to QUD: Alternatives to questions. *Functions of Language* 29(1). 86–115. DOI: 10.1075/fol.00039.oze.

Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27(1). 53–94. DOI: 10.21825/philosophica.82606.

Renans, Agata, Malte Zimmermann & Markus Greif. 2010. *Questionnaire on focus semantics* (Interdisciplinary Studies on Information Structure: ISIS; Working Papers of the SFB 632-15). Potsdam: Universitätsverlag Potsdam. https://publishup.uni-potsdam.de/frontdoor/index/index/docId/4714.

Repp, Sophie. 2010. Defining "contrast" as an information-structural notion in grammar. *Lingua* 120(6). 1333–1345. DOI: 10.1016/j.lingua.2009.04.006.

Rooth, Mats Edward. 1985. *Association with focus*. Amherst, MA: University of Massachussetts. (Doctoral dissertation).

Rooth, Mats Edward. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116. DOI: 10.1007/BF02342617.

Rooth, Mats Edward. 1996. On the interface principles for intonational focus. *Semantics and Linguistic Theory* 6. 202–226.

Sasse, Hans-Jürgen. 1987. The thetic/categorical distinction revisited. *Linguistics* 25(3). 511–580.

Sasse, Hans-Jürgen. 1996. *Theticity*. Working paper 27. Cologne: Institut für Sprachwissenschaft der Universität zu Köln.

Sasse, Hans-Jürgen. 2006. Theticity. In Giuliano Bernini & Marcia L. Schwarz (eds.), *Pragmatic organization of discourse in the languages of Europe*, 255–308. Berlin: Mouton de Gruyter.

Sheehan, Michelle & Jenneke van der Wal. 2018. Nominal licensing in caseless languages. *Journal of Linguistics* 54(3). 527–589. DOI: 10.1017/S0022226718000178.

Sikuku, Justine & Michael Diercks. 2021. Object marking in Lubukusu: Information structure in the verb phrase. In Akinbiyi Akinlabi, Lee Bickmore, Michael C. Cahill, Michael Diercks, Laura J. Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 Years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 305–330. Berlin: Language Science Press. DOI: 10.5281/zenodo.5578822.

Šimík, Radek. 2018. Ever free relatives crosslinguistically. *Proceedings of Sinn und Bedeutung* 22(2). 375–392.

Skopeteas, Stavros, Ines Fiedler, Samantha Hellmuth, Anne Schwarz, Ruben Stoel, Gisbert Fanselow, Caroline Féry & Manfred Krifka. 2006. *Questionnaire on Information Structure QUIS* (Interdisciplinary Studies on Information Structure 4). Potsdam: University of Potsdam.

Szabolcsi, Anna. 1981. Compositionality in focus. *Folia Linguistica* 15(1-2). 141–162. DOI: 10.1515/flin.1981.15.1-2.141.

Van de Velde, Mark. 2019. Nominal morphology and syntax. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 2nd edn., 237–269. London: Taylor and Francis.

van der Wal, Jenneke. 2009. *Word order and information structure in Makhuwa-Enahara*. Leiden University. (Doctoral dissertation).

van der Wal, Jenneke. 2011. Focus excluding alternatives: Conjoint/disjoint marking in Makhuwa. *Lingua* 121(11). 1734–1750. DOI: 10.1016/j.lingua.2010.10.013.

van der Wal, Jenneke. 2015. Bantu syntax. In *Oxford handbook Topics in linguistics*. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199935345.013. 50.

van der Wal, Jenneke. 2016. Diagnosing focus. *Studies in Language* 40(2). 259–301. DOI: 10.1075/sl.40.2.01van.

van der Wal, Jenneke. 2021. On the use of idioms for testing focus. *Semantic Field Methods* 3(2).

van der Wal, Jenneke & Amani Lusekelo. 2022. The V and CV augment in Kinyakyusa and exhaustivity. *Studies in African Linguistics* 51(2). 324–346.

van der Wal, Jenneke & Jacky Maniacky. 2015. How 'person' got into focus: Grammaticalization of clefts in Lingala and Kikongo areas. *Linguistics* 53(1). 1–52. DOI: 10.1515/ling-2014-0033.

van der Wal, Jenneke & Saudah Namyalo. 2016. The interaction of two focus marking strategies in Luganda. In Doris L. Payne, Sara Pacchiarotti & Mokaya Bosire (eds.), *Diversity in African languages*, 355–377. Berlin: Language Science Press.

van der Wal, Jenneke & Aurélio Simango. 2022. Subject inversion in Changana. In Ezra Chambal Nhampoca & Letícia Cao Ponso (eds.), *Pesquisa e ensino em línguas moçambicanas. Um tributo a Bento Sitoe*, 229–247. Maputo: Gala-Gala Edições.

Vermeulen, Reiko. 2012. Word order variation and information structure in Japanese and Korean. In Ad Neeleman & Reiko Vermeulen (eds.), *The syntax of topic, focus, and contrast*, 77–118. Berlin: De Gruyter Mouton. DOI: 10.1515/9781614511458.77.

Yoneda, Nobuko. 2011. Word order in Matengo (N13): Topicality and informational roles. *Lingua* 121(5). 754–771. DOI: 10.1016/j.lingua.2010.11.008.

Zimmermann, Malte. 2016. Predicate focus. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure*, 314–335. Oxford: Oxford University Press.

Zimmermann, Malte & Edgar Onea. 2011. Focus marking and focus interpretation. *Lingua* 121(11). 1651–1670.