

In search of larger units of meaning: a foray into Northern Sotho dataⁱ

Elsabé Taljard, Department of African Languages, University of Pretoria, e-mail: elsabe.taljard@up.ac.za

Abstract

The availability of large scale electronic language corpora has led to revolutionary changes in the landscape of linguistic research. Corpus linguistics has in the past few decades developed from an alternative methodology of linguistic investigation to a fully-fledged theoretical approach to linguistic study. Furthermore, having access to electronic corpora has lead scholars such as Sinclair (2004) to redefine the basic unit of meaning of language. In the light of corpus evidence, he proposes the existence of a lexical unit which represents a lexical structure higher than the word, in which meaning is vested. In this article, the notion of a functional unit of meaning in Northern Sotho is explored. Two case studies are presented, illustrating not only the methodology of corpus-driven investigation for a Bantu language, but also the theoretical and practical implications of recognizing the existence of extended units of meaning.

Keywords

extended unit of meaning, collocation, colligation, semantic prosody, Northern Sotho, corpus-driven language study, Bantu linguistics

From corpus-based to corpus-driven linguistic investigation

The availability of large scale electronic language corpora has led to revolutionary changes in the landscape of linguistic research. Initially, corpus linguistics was merely seen as an alternative methodology to the study of language, where the role of the corpus is to provide easy access to a large amount of data, to a repository of examples that can be used to support a linguistic argument or to test and / or validate some theoretical statement. This is known as a corpus-based approach to linguistic investigation. Corpus-based linguistic study is always done within a particular

theoretical framework, often within generally accepted dichotomies such as competence / performance, lexis / grammar; evidence gleaned from a corpus is used to validate existing categories and at best, adds a probabilistic dimension to a pre-existing theory. Tognini-Bonelli (2001,10) however argues that ‘in this context, the potential of corpus evidence is not exploited fully because, in order not to threaten dramatically some existing theoretical positions, the richness of language usage is in many ways sacrificed and it is not allowed to shape the descriptive and theoretical statements that should ideally account for it.’ In a corpus-driven approach, the corpus is seen as much more than a source of examples to support some pre-existing theory. Here, the starting point is the evidence produced by the corpus, and based on this evidence, a theory or at least some theoretical statement needs to be formulated to account for the data provided by the corpus. Sinclair (2004, 10) advises the corpus linguist to ‘inspect the data with as little attention as possible to theory’. Actual language use therefore informs the formulation of linguistic theory – an approach that is in sharp contrast with the Chomskyan tradition, in which language use is deemed to be of no importance to the formulation of linguistic theory; intuition and introspection are the mainstays of linguistic investigation. It is therefore clear that corpus-driven linguistic investigation is more than simply a new methodology, since it impacts directly on the theoretical position of the linguist. In contrast to corpus-based linguistic studies, a corpus-driven approach does not function within an existing framework, but in many cases challenges these frameworks and explores whether they are indeed borne out by corpus evidence.

The change in the relationship between language data and linguistic theory is not the only one brought about by a corpus-driven approach to the description of language. The pre-theoretical analysis of corpus evidence has the potential to reveal the existence of meaningful patterns that may go unnoticed, not only by the traditional linguist, but even to a linguistically aware mother tongue speaker. It is only when confronted by a page or pages of concordance lines that the linguist can recognize these patterns and account for them within the description of a grammar.

A third issue raised by a corpus-driven approach is the interconnection between an item and its environment. According to Tognini-Bonelli (2001, 101), the meaning of a word is determined by the formal co-textual features surrounding it, and there is thus

no separation between an item and its environment. Central to this premise is the notion of co-selection, i.e. the simultaneous selection of two or more lexical items, resulting in the establishment of an extended unit of meaning. Sinclair (2004, 171) uses the phrase ‘out of the corner of my eye’ to argue convincingly that the seven words constituting this phrase, represent a single, simultaneous choice having to do with peripheral vision. The inseparability of item and environment and the concept of co-selection culminate in the formulation of the idiom principle, which is returned to below.

It is exactly the availability of large scale electronic corpora that leads Sinclair (2004, 24 et seq.) to redefine the basic unit of meaning of language. In the light of corpus evidence, he proposes the existence of a lexical unit which represents a lexical structure higher than the word, in which meaning is vested. Tognini-Bonelli (2001, 11) refers to this type of unit as a functionally complete unit, one which has reached its semantic prosody. She uses the adjective ‘proper’ to illustrate the notion of a functionally complete unit. By analysing the collocational profile of this word, she is able to identify an extended unit of meaning within which the word ‘proper’ functions. She finds that this extended unit of meaning does not coincide at all with the assumed meaning of ‘proper’ as ‘appropriate’. The choice of the word ‘proper’ is therefore a functional one, which goes beyond labelling something as ‘appropriate’ – ‘the function of *proper*, therefore, finds its place in an extended unit of meaning which is “a complaint for the absence of something that we all think should be present or available”’.

The aim of this article is to explore the notion of functional units of meaning in Northern Sotho. The reader could rightly ask whether such an endeavour is worth the trouble and whether it would not simply imply a projection of the findings for English onto the Northern Sotho data, but Sinclair (2004, 19) makes provision for the possibility that languages with different principles of word construction may present a different perspective on issues such as collocation, which is a notion central to the establishment of larger units of meaning. This in itself provides adequate motivation for this study: like all Bantu languages, Northern Sotho is a predominantly agglutinating language, making extensive use of affixes for word formation, whereas a language such as English is mainly analytic. Should it be found that no significant

adaptation to the theoretical underpinnings of the notion of larger units of meaning, based on what the Northern Sotho data reveal, is necessary, this could imply that the morphological typology of a language does not affect the way in which meaning is assigned to larger lexical chunks. In this way, the findings would contribute to the existing body of knowledge regarding larger units of meaning, and would thus address one of the outstanding theoretical issues. This investigation thus represents a pioneering corpus-driven study of larger units of meaning for Northern Sotho in terms of semantic prosody, collocations and colligation. It is the first investigation of its kind, not only for Northern Sotho, but also for any of the South African Bantu languages.

The proposed investigation depends on the availability of a sizeable electronic data corpus which enables the linguist to process and systematically examine a vast quantity of Northern Sotho data with the aid of computational tools. The availability of a 7,2 million word corpus for Northern Sotho, the University of Pretoria *Sepedi Corpus* (PSC), makes such an endeavour possible.

Theoretical conspectus

One of the intrinsic principles upon which language description is based, is the assumption that the word is the basic unit of meaning. Sinclair (2004, 24) calls the word one of the primitives of language, the other being the sentence. Taking the word as the primary unit of lexical meaning furthermore implies that a word has an independent meaning. The way in which words are treated in dictionaries illustrates this mode of thinking rather clearly. Compare the entry for the word *hlahla* from the Bilingual Northern Sotho – English dictionary (De Schryver, 2007):

hlahla verb 1 ► **coach** Thellenyane o kgethilwe go hlahla sehlopha sa kgwele ya maoto sa sekolo. • *Thellenyane was selected to coach the school soccer team.* 2 ► **conduct (a choir)** Se se ra gore o tlo hlahla le khwaere ya sekolo. • *This means he will also conduct the school choir.* 3 ► **lead** Baetapele ba ile ba hlahla mogobo go ya go fihla seferong sa kantoro ya balaodi. • *The leaders led the demonstrators until they reached the entrance to the management's office.*

From this example it is clear that words are regarded as the basic units of meaning; that they can have several meanings, which are then listed separately in a dictionary. However, there are instances where the independence of the word as unit of meaning is compromised, idioms, proverbs and fixed phrases being the most well-known and oft-cited examples. In conventional grammatical descriptions of a language, these extended units of meaning are largely ignored, being regarded as somewhat quirky vocabulary items that have to be learnt, since the meaning of such a multiword unit is more than the combined meaning of its parts. This is also true with regard to Northern Sotho, where learners of the language are invariably confronted with lists of ‘idiomatic expressions’ which have to be learnt off by heart, since they do not fit into any descriptive framework. Sinclair (2004, 28) further points out that the composite parts of these units of extended meaning appear in a variety of relationships to the meaning of the unit. In some cases none of the words seems to contribute to the meaning of the expression, e.g. (*go*) *hlaba thedi*, which can roughly be translated as ‘(to) curtsy’, where neither the meaning of *hlaba* (‘stab; slaughter’), nor the meaning of *thedi* (‘haunches’) contributes to the meaning of the whole. In *swerwe ke tlala* (‘be hungry’) the meaning of *tlala* (‘hunger’) has been retained, but not that of *swerwe* (‘be held’).

Based on evidence provided by the corpus, Sinclair (1996, 2004) argues that the phenomenon of extended units of meaning is not restricted to idioms, proverbs and fixed expressions, and that the choice of a word is very rarely independent of the environment in which it appears. This view is echoed by Hanks (2006, 19), who indicates that words only have meanings when they are put into context. When in isolation, words only display meaning potential, which consists of ‘any number of rather fuzzy semantic components, some or all of which are activated when the word is used’. Sinclair consequently distinguishes between the phraseological tendency, which he defines as the tendency of the speaker / writer to choose several words at the same time, and the terminological tendency, which refers to the tendency of a word to have a fixed meaning, so that the meaning it delivers is guaranteed every time it is used. Furthermore, any person wanting to refer to the referent of such a word has little choice but to use it. As the name suggests, the terminological tendency applies mainly to items of a terminological nature, where the relationship between the term and its

referent is fixed. In cases such as these, words would be selected according to what Sinclair (2004, 30) calls the open-choice principle, which implies that words have a meaning independent of their environment. However, Sinclair (1991, 110) indicates that the open-choice principle is inadequate to account for meaning in language in that ‘the open-choice principle does not provide for substantial enough restraints on consecutive choices’. Since operation of the open-choice principle alone would not enable language users to produce normal text, the principle of idiom is put forward to account for the restraints that cannot be captured by the open-choice model. The idiom principle, also referred to as the phraseological tendency, implies the simultaneous choice of two or more words, resulting in what Francis (1993, 142) terms a ‘single-choice chunk’ and constitutes a multi-word lexical item or an extended unit of meaning, where the boundaries of meaning encompass the unit as a whole. A simple example would be the phrase *bjalo ka* (‘such as’), which operates as a single word, even though it seems to be analysable into two segments. Sinclair (1991, 110) refers to the space between these two elements as ‘structurally bogus’, which only serves to mask the fact that it represents a single choice. Also compare the following, slightly more complex example, used here to illustrate the features exhibited by the idiom principle. There is a phrase in Northern Sotho *-tomoletše motho mahlo*, which can loosely be paraphrased as ‘stare someone (or less frequently something) straight in the eye, sometimes in fear, anger or surprise’. In actual language usage as revealed by the corpus, this semi-preconstructed phrase typically consists of a subject, often represented by a subjectival concord functioning as a pronoun, the bitransitive verb stem *-tomoletše* (verb root *tomol-* + applied extension *-el-* + perfect tense suffix **-ile*), an indirect object (IO) and the noun *mahlo* (‘eyes’) as direct object (DO). These different orthographic units represent a case of co-selection, where more than one word is selected in a single choice, which has to do with (unblinking) staring. Sinclair (1991, 111) indicates that these phrases are often of indeterminate extent and it is therefore not always possible to determine which elements form an integral part of the phrase, and which ones are the results of collocational attraction. With regard to the current example, this is a moot point, since *-tomoletše* is a verb stem and with the exception of verbs appearing in the imperative and infinitive moods, the presence of a subject concord is an obligatory feature of all verb stems in Northern Sotho. It can therefore safely be assumed that whatever subject is present forms an integral part of this particular phrase. A second feature of the idiom principle is that some elements

within the phrase are variable: the subject and consequently the subject concord (sc) are variable (cf. line 22: subject *Tsietsi* ('Tsietsi'), sc *a*, line 26: subject NP *batswadi bona le Thanthankedi* ('the parents and Thanthankedi'), sc *ba*). The same goes for the indirect object (cf. line 25: IO *monna yo a rego ke Mokowe* ('a man called Mokowe'), line 26: IO *Ngaka Mošunkutšwane* ('Doctor Mošunkutšwane') and the object concord (oc) (cf. line 22: oc *n-* ('me'), line 23: oc *re* ('us'), line 28: *ba* ('them')). These variables constitute internal lexical variation. Some syntactic variation can also be present in this phrase: the indirect object can either appear as a noun or noun phrase in the immediate postverbal slot (cf. lines 25 and 26), or it can be pronominalized and be represented by an objectival concord in the position immediately preceding the verb stem, cf. lines 23, 24, 28, 29. Some variation in word order is also allowed in some phrases, cf. line 27, in which the indirect object *Thušano* ('co-operation') appears in the preverbal position instead of in its basic postverbal position, immediately following the verb stem. Lexical variation, syntactic variation and variation in word order constitute the full extent of variation in this phrase; the remaining words, i.e. the verb stem *-tomoletše* and the direct object *mahlo* ('eyes') are fixed.

Figure 1: Excerpt from KWIC lines for *tomoletše

N	Concordance
22	tla apara eng ge re di hlatswitše?" Tsietsi o mpotšiša a tomoletše mahlo. "Re tla apara eng? Ge re se na se
23	yo mobjang. Monna yo a bego a eme lebating a re tomoletše mahlo, o be a se a apara hempe goba dieta.
24	ka gore šefa eke ge e le maetwaetwana ao a bego a mo tomoletše mahlo o a phethile. Afaeya, a ka lehutša ka
25	lena šomelang ya lena tšhelete." Sebaka o bolela bjalo a tomoletše monna yo a rego ke Mokowe mahlo. Ke a
26	ba gagwe. Batswadi bona le Thanthakedi ba be ba tomoletše Ngaka Mošukutšwane mahlo ba re a tšwe ka
27	re tswiotswio Thušo re e nyaka ka moka, Thušano re e tomoletše mahlo ka moka, Thulano re e hloile ka moka,
28	gare ka lefastere. Letl. 45). Le ge kotsi re e bona e ba tomoletše mahlo ga ba iše felo ka yona. Ba bona
29	realo a phološa Seitshwenyeng lehung leo le bego le mo tomoletše mahlo. Seitshwenyeng o ile a gola bjalo ka

It can therefore be concluded that the phrase consisting of a variable subject + *-tomoletše* + variable indirect object + *mahlo* constitutes an extended unit of meaning, since the unit starts with a node as core, i.e. *-tomoletše* and *mahlo*, then incorporates other words in the co-text that seem to be simultaneously selected with it and eventually results in a regular pattern. In Tognini-Bonelli's (2001, 19) words, such multi-word units are defined by the strict correlation which exists between a node and its context.

Sinclair (2004, 30) argues that within these extended units of meaning, the interconnection between an item and its environment is so strong that the environment

actually becomes part of a multi-word lexical unit: ‘... words cannot remain perpetually independent in their patterning [...] they begin to retain traces of repeated events in their usage, and expectation of events such as collocations arise’. In these extended units, the independence of the choice of words can be influenced by both lexico-grammatical and semantic constraints.

As was suggested above, the notion of an extended unit of meaning is best illustrated with reference to idioms and other fixed expressions, but it seems that this phenomenon is much more pervasive than these formulaic expressions. It is possible for any single word to become so closely integrated with its cotext that its cotextual features determine its meaning, resulting in the inseparability of meaning and environment.

Sinclair (2004) describes the notion of extended units of meaning in terms of four interrelated concepts, which represent different levels of abstraction. These concepts are collocation, colligation, semantic preference and semantic prosody. As pointed out by Nelson (2000, 181), collocation and semantic prosody are both concerned with the lexical patterning of words, whereas colligation is concerned with the grammatical patterning of words. Semantic preference controls the collocational and colligational patterns. It is clear that these concepts are not totally separate, but that they are interrelated and interdependent, and that together they create a network of meaning. More detailed descriptions of these concepts will follow in the discussion below.

Extended units of meaning: two case studies

The first case study presented here concerns the verb (*go*) *kgotlelela* (‘(to) tolerate, persevere’). In order to identify regular, significant patterns – be they lexical, grammatical or semantic – associated with the use of the verb stem *-kgotlelela* (‘tolerate, persevere’), the 7,2 million word PSC (University of Pretoria *Sepedi Corpus*) was queried. Due to the disjunctive writing tradition followed in Northern Sotho, the verb stem is used as the search node, since the various verbal prefixes which can potentially precede the stem are variable and represent an almost infinite number of possible combinations.

Using WordSmith Tools' concordancing function, a total of 288 KWIC lines (keyword in context) was thrown up for the form *-kgotlelela*. According to the English - Northern Sotho School Dictionary (De Schryver, 2007), the verb stem *-kgotlelela* has two senses, i.e. 'persevere' and 'tolerate'. This distinction is also borne out by the concordance lines, but more significant is the fact that these two senses correlate roughly with distinct grammatical patterns, confirming the notion that 'every sense or meaning of a word has its own grammar' (Francis 1991, 145), a point which will be returned to below. In the case of the verbs containing the stem *-kgotlelela*, perusal of the concordance lines indicates that the meaning '(to) persevere' is linked to the intransitive use of the verb, whereas the meaning '(to) tolerate' is realised whenever the verb is used transitively, or followed by an infinitive clause. Compare the following excerpts from the concordance lines culled from the corpus. Note that translations of the relevant sections have been inserted for illustrative purposes:

Figure 2: KWIC lines for *kgotlelela* ('persevere')

- 1 go direla setšhaba. Meladi o a **kgotlelela** o sa leka gape go
to serve the nation. Meladi perseveres, still trying to
- 2 a nego ngwanešo. Sa gago ke go **kgotlelela.**" "Ke bona ke tla
Yours is to persevere. I see that I'll
- 3 o kgotlelela." "Ke bona ke tla **kgotlelela** ka lebitleng, e se
I see that I will persevere in the grave
- 4 tso e mmalwa a dutše a re ke a **kgotlelela** fela ka morago o
many times he continued to persevere but later he
- 5 ye a lekago go e bala, a se sa **kgotlelela.** "Le a re foka man
which she tried to read, she no longer persevered. You are
- 6 Taamane mo a se sa kgonago go **kgotlelela.** Le gona bommagwe
where Taamane could no longer persevere. Furthermore, her mother
- 7 seo se setelele Maite a se sa **kgotlelela,** "Bjale ge go le ka
Maite could no longer persevere, "Now if it is
- 8 gago. Ke bile ke feditše e dio **kgotlelela.** (O feditše.) Ke g
I have finished, just persevere. (He is has finished)
- 9 we a ka ba mahlatse ge a ka no **kgotlelela.** A - Re tla re ke
he will be fortunate if he can only persevere. We'll say it's
- 10 bjalo a hwetše bokaone e le go **kgotlelela** mošomong gore a
she found it would be better to persevere in the job so that

Figure 3: KWIC lines for *kgotlelela* ('tolerate')

1	batho ba bantši ba sa ba kgona go	kgotlelela	mathata. Monna yola <i>many people cannot tolerate difficulties</i>
2	Mongangale. Taba ye nka se sa e	kgotlelela,	e ntapišitše." Le <i>Mongangale. This issue I can no longer tolerate, it has tired me out</i>
3	ke motho wa go se kgone go	kgotlelela	tlala. Ke bone gore <i>is a person who cannot tolerate hunger</i>
4	na? Efela Gabantsebe o ile a bo	kgotlelela,	a bo dulela go fihla <i>But Gabantsebe tolerated it, he sat it out until</i>
5	magareng ga monna le mosadi, Le	kgotlelela	madimo le diphefo <i>between husband and wife you tolerate storms and wind</i>
6	hlagiše dibjalo tšeo di kgonago go	kgotlelela	komelelo goba malwetši <i>grow plants which can tolerate drought or disease</i>
7	maaka ke selo se ke sa kego ke se	kgotlelela	le gannyane." Ke Ariel <i>lies are something which I cannot tolerate even a little bit</i>
8	ka lebaka la eng re swanetše go	kgotlelela	baagišane ba mohuta <i>why we must tolerate this kind of neighbours</i>
9	tša Afrika - Borwo di kgona go	kgotlelela	komelelo le malwetši <i>of South Africa can tolerate drought and disease</i>
10	hwetša mehlare yeo e kgonago go	kgotlelela	komelelo, bjalo ka <i>find trees which can tolerate drought such as</i>

For the purpose of this discussion, it was decided to focus upon incidences such as those in Figure 3, where the verb containing the stem *-kgotlelela* is used transitively, with the meaning of 'tolerate'. The investigation is further narrowed down to include only those cases in which the object NP is explicitly stated, thus excluding examples where the object NP has been pronominalized, either by means of an object concord preceding the verb stem, or by means of any other pronominal expression occupying the slot immediately following the verb. The initial 288 concordance lines were then manually perused in order to identify the relevant ones, i.e. those concordance lines in which *-kgotlelela* is used to express the sense of 'tolerate'; secondly, to eliminate obvious 'noise' produced by the corpus search and thirdly, to select concordance lines in which the object NP is explicitly stated. This manual selection and cleaning process resulted in 67 KWIC lines being available for analysis; these are listed in appendix A.

Following the methodology set out in Sinclair (2004, 30 et seq.) and Tognini-Bonelli (2001, 106), the analysis of the company that the verb stem *-kgotlelela* keeps, is done in terms of the four key concepts mentioned above, starting with the collocational

properties of *-kgotlelela*. Nelson (2000) cites no less than 9 different definitions for the term collocation, illustrating that the term is used and understood in many different ways. Sinclair (2000, 200) defines it simply as the co-occurrence of words, which seems to capture the essence of the concept, i.e. that words keep company with each other. Smadja (1993, 143) adds a statistical dimension by stating that collocations are recurrent combinations of words that co-occur more often than expected by chance and that correspond to arbitrary word uses. This implies that the collocational relationship between an item and its company can be statistically determined. Furthermore, from the analysis of examples provided by different scholars, it is clear that the words making up a collocation need not appear adjacent to each other, but can be separated by a varying number of other words.

Looking at object NPs that typically co-occur with verbs containing the verb stem *-kgotlelela*, some elements of co-selection can be noted at the collocational level: two nouns that recurrently occur are *mathata*, (with its variant *mabothata* and its singular form *bothata*) ('trouble, problem(s)') and *komelelo* ('drought'). Together, these items cover 25% of object NPs co-occurring with *-kgotlelela*. What is also evident is that the other object NPs which co-occur with this verb all share the semantic implication of hardship, cf. *ditlaišego*, *ditlaišo* ('suffering'), *tlala* ('hunger'), *maroga* ('cursing'), *botšididi* ('cold'), *mengunanguna ya mmagwe* ('his/her mother's grumbling'), *phišo* ('heat'), *malwetši* ('diseases'), *ditlhokofatšo* ('hardships'), *bohloko* ('pain'), etc. It is therefore clear that whatever needs to be tolerated is something unpleasant. Even humans that have to be tolerated have negative characteristics: an analysis of the discourse context from which KWIC line 1 has been taken reveals that the neighbours (*baagišane*) which are being referred to are typical problem neighbours, and that this specific concordance line appears in a letter of complaint about the errant neighbours. In KWIC line 34, *mašilo* ('fools') have to be tolerated; in line 51, a woman has to tolerate a man that does not eat at home at night: *monna wa go se je ka gae mantšiboa*, the implication being that he is sexually unfaithful – undoubtedly a negative trait. The verb stem *-kgotlelela* therefore clearly collocates with a semantic set consisting of a group of words related to unpleasantness, hardship and suffering. At this point, the analysis starts to reveal the semantic prosody associated with the verb stem *-kgotlelela*, which Sinclair (1998, 20) defines as the reason why an item is chosen over and above the semantic preferences that also characterize it. He maintains

that semantic prosody is something that reflects the attitude of the speaker, and is to be found on the pragmatic side of the semantics/pragmatics continuum. Louw (2000, 9) refines the definition by stating that ‘a semantic prosody refers to a form of meaning which is established through the proximity of a consistent series of collocates, often characterisable as positive or negative, and whose primary function is the expression of the attitude of its speaker or writer towards some pragmatic situation’.

It would seem then, that the dividing line between the verb containing the stem *-kgotlelela* and its environment is somewhat blurred, that the element of hardship present in the surrounding context has somehow fused with the meaning of the verb stem. By selecting the verb stem *-kgotlelela*, the speaker has already committed him/herself to selecting an object referring to something which is unpleasant as an adjunct to the verb – a notion which ties in with Sinclair’s (2004, 19) remark that collocation can be described by saying ‘that the choice of one word conditions the choice of the next, and of the next again’. The decision to use this particular verb stem leads to more than one word in the text. The semantic preference identifies the semantic field within which the verb stem *-kgotlelela* operates. Thus, the selection of the verb and the object which co-occurs with it presents a single, simultaneous choice on the part of the speaker.

Further clues to the semantic prosody of the stem *-kgotlelela* (‘tolerate’) is to be found in the left co-text of the verb. We find that the object NPs that typically co-occur with *-kgotlelela* are difficult or even impossible to tolerate, as evidenced by the often negative polarity revealed by an analysis of the co-text to the left of the verb containing this stem. From the KWIC lines, it is evident that *-kgotlelela* often appears as an infinitive adjunct to the verb stem *-kgona* – a modal verb stem expressing a modality of possibility. The same modality is also expressed by means of the potential morpheme *-ka*, and together these two items cover 28% of all occurrences of *-kgotlelela*. What is further significant is that these items often appear in a negative construction, thus indicating a modality of impossibility. Compare *ga e sa kgona go kgotlelela* (‘it can no longer tolerate’) (line 9), *se ka se sa kgona go kgotlelela* (‘it can no longer tolerate’) (line 17 & 29), *ba sa ba kgona go kgotlelela* (‘(if) they cannot tolerate’) (line 42), *ga ke sa kgotlelela* (‘I no longer tolerate’) (line 60), etc. In some

examples the notion of impossibility is realized on the lexical level, cf. *a šitwa ke go kgotlelela* ('he/she is unable to tolerate') (line 33 & 47), *o ile a palelwa ke go kgotlelela* ('he/she was unable to tolerate') (line 16), *ga se thaka ya mošemane go kgotlelela* ('it is not child's play to tolerate') (line 4). Tolerance of hardships therefore does not come easily. In the few examples where *-kgotlelela* is not preceded by a negative verb, we see that tolerance is something that needs to be taught, that needs to be assisted, as evidenced by the occurrence of the verb stems *-kgontšha* ('enable'), and *-thuša* ('help'). An element of obligation is found in the use of the auxiliary *-swanetše* ('must') and the verb stem *-tlamegile* ('be compelled'). The co-text preceding the stem *-kgotlelela* therefore clearly expresses a semantic prosody of something that we ought to be able to do, that is possible to do, but that we more often than not, find difficult and/or impossible.

Having reached the semantic prosody, we have also reached the boundaries of the extended unit of meaning, which can be paraphrased as follows: The speaker selects a semantic prosody of difficulty/impossibility, linked to a semantic preference of unpleasantness. The semantic preference controls the collocational pattern of object NPs selected as adjuncts of the verb containing the stem *-kgotlelela*. We can therefore conclude that this stem is the core of a functionally complete unit of meaning. The selection of the verb stem *-kgotlelela* by a speaker therefore far surpasses the notion of tolerance; it actually expresses the inability and/or difficulty to tolerate something that is invariably unpleasant. We have thus arrived at a model of a single lexical item consisting of several (linguistic) words, which can be represented as follows: difficulty / impossibility + *-kgotlelela* + unpleasant object.

As a second case study, a lexical item in which the verb stem *-swerwe* appears is investigated. This verb stem is a derived form of the inchoative stem *-swara* with the canonical meaning 'get hold of, grab, take' to which a perfect suffix *-ile* and a passive verbal extension have been added. Perusal of the 603 concordance lines thrown up by a corpus search reveals what seem to be three distinct patterns in which *-swerwe* appears. Compare the following examples:

Figure 4: Pattern 1 for –*swerwe*

N	Concordance
328	a yo nyala Mafodisa. Badimo ba ganne, gomme go <i>swerwe</i> yena. Go napile go senyegile, morago molato o
329	ba hlakahlakane le bafelegetši ba bona, ka moka go <i>swerwe</i> dikgati tšela magadi ale a go ikotla ka tšona. Ke
330	ke mantšu ao a felago a tshela gare nako le nako ge go <i>swerwe</i> mehlamu. Gapeletšang gore go be ka mokgwa
331	tša letšatši. Kgobokanong ka segotlong go be go <i>swerwe</i> mehlamo ya fasana ka ge go felwa pelo ya go
332	ya ka lapeng la modula setulo wa komiti ba tla hwetša go <i>swerwe</i> Lekota. Ditaba tše bjalo e be e se maaka ka
333	Phethiši pele go feta matšatši a 7 go tlogela ge go <i>swerwe</i> kopano ya Khansele ya go lekola mongwalo wa
334	. Tšatši le a go nyaka go hlahlelwa o be a hweditše go <i>swerwe</i> tsotsi ye nngwe e gana go fetola dipotšišo. Ba

In this pattern the subject and agent of the verb is unspecified, as is evidenced by the use of the indefinite subject concord *go-* preceding the verb stem. The patient appears in the post verbal position and represents mostly new or indefinite information which is introduced into the discourse for the first time. In order to get a sense of the meaning conveyed by this construction, compare an analysis of KWIC lines 331 and 334 by way of illustration:

(1) KWIC line 331

<i>go</i>	<i>be</i>	<i>go</i>	<i>swerwe</i>	<i>mehlamo</i>	<i>ya</i>	<i>fasana</i>
SC Indef	AUX	SC Indef	Vst-*il-w-e	N4	PC4	N16-dim
there	(was)	there	is held	conversation	of	low
‘a whispered conversation was held’						

KWIC line 334

<i>go</i>	<i>swerwe</i>	<i>tsotsi</i>	<i>ye</i>	<i>nngwe</i>
SC Indef	Vst-*il-w-e	N9	QUAL PART9	ADJ
there	was caught	tsotsi	-	another
‘another tsotsi was caught’				

Figure 5: Pattern 2 for -*swerwe*

N	Concordance
98	Kwano ya Leilane le Maphuthe e <i>swerwe</i> ke matšhona, e bile le go tshepana ga
99	. Tše dingwe di hwile dithunthwane o ka re di <i>swerwe</i> ke malopo. Mosadi wa batho o ile go
100	gona? Ga ke kgolwe!" "Šedio ge!" "Ba re o <i>swerwe</i> ke bolwetši bofe?" "Fao gona ke
101	. Morena Matlala ka nnete ke kgale a <i>swerwe</i> ke tlala ya go tsokama setulo sa
102	ke bolwetši bja motšatši. Gonabjale bana ba <i>swerwe</i> ke kwhaši, dimpana ke mararampana
103	dijo, ba bile ba go ganetša le tšona dijo mola o <i>swerwe</i> ke tlala ba re wena o direng? Re ile ra
104	e le gore go be le madi go bontšha gore o <i>swerwe</i> ke maphodisa. Seo se dirwa ka nepo

Figure 6: Pattern 3 for –*swerwe*

N	Concordance
10	. Ke tla dira bjang ka gore go swarwa gona ke <i>swerwe</i> , a nape a ineela matsogong a bona ka
11	: ke letela bona. E tlo ba dihlong tša hodu ge le <i>swerwe</i> . A ba lewe ke dihlong, ba moloko wa
12	ka molato wa Themba wa ngwagolola ge a be a <i>swerwe</i> ba re o nyakile go bolaya ralebenkele
13	gare ga ba fedilego ka marumo. Lerumo le <i>swerwe</i> ; ba aogeng le mabotho ohle a bona.
14	gore ba mo swere. Le yena o lemoga gore o <i>swerwe</i> . Ba a tseba gore dilo tše pedi tšeo:
15	ke batho ba go se tshephiše mo gongwe ba <i>swerwe</i> ba dirile bohvirihwiri. Ebile go feta fao
16	no gana go kwa gore basenyi ba bangwe ba <i>swerwe</i> ba nape ba tsupuloga ba re le rena ba

At a first glance, the KWIC lines in figures 5 and 6 seem to represent different patterns. These two patterns are however, similar in all respects, except for the (non)specification of the agent. In figure 5, the agent is specified and appears in an agentive phrase, following the agentive prefix *ke-*. In figure 6 the agent is left unspecified, since it can probably be inferred from the context or is deemed to be redundant from a discourse point of view. In both patterns, the patient is present in the left context of the verb stem, either as a noun, in which case the subject concord following it functions as a marker of the grammatical subject (cf. figure 5, line 102 *bana ba* ('children') and figure 6, line 13 *lerumo le* ('spear')), or in pronominal form, in which case the subject concord functions as a pronoun, representing the subject NP in the verb. The rest of the discussion will focus mainly on the pattern illustrated in Figure 5, which is traditionally regarded as the prototypical passive structure.

Further analysis reveals that in cases where the agent NP refers to an animate object, one of several, mostly literal, meanings is expressed. Cf. the following examples culled from the corpus:

(2) 'be arrested'

... *ka tšhoga go lemoga gore ke swerwe ke sersanta Motle*

('I was shocked to realize that I am being arrested by sergeant Motle')

(3) 'be caught'

O swerwe ke dimpša tša rena.

('He is caught by our dogs.')

(4) 'be held'

Seo se bego se swerwe ke Faro e be e le sethunyana

‘That what was held by Faro was a small gun’

(5) ‘be touched’

... ka go re e swerwe ke diatla tše ntši nakong ya ge e šilwa

(‘... because it is touched by many hands at the time when it is minced’)

(6) ‘be taken’

go lahla tsela ye e bego e swerwe ke botatabo

(‘to abandon the road that was taken by their fathers’)

(7) ‘be handled’

Taolo ya SITA e swerwe ke Boto ya Balaodi

(‘The management of SITA is (being) handled by a Board of Directors’)

It could be argued that the agent in (5) above (*diatla* ‘hands’) does not refer to an animate object, but this is clearly a metonymical expression, where ‘hands’ refer to humans, of which hands are a component part. The same applies to (7).

In examples where the agent NP refers to an inanimate object, the meaning of the verb containing the verb stem *-swerwe* is of a non-literal / figurative / metaphorical nature, i.e. ‘be afflicted, troubled’. Compare the following examples:

(8) *Morwedi wa monna yoo o be a swerwe ke bolwetši bjo šoro kudu.*

(‘The daughter of that man was afflicted by a very serious disease.’)

(9) *Phiri e be e swerwe ke tlala*

(‘The hyena was troubled by hunger, i.e. was hungry’)

(10) *Bona ba swerwe ke tšhogo e kgolo*

(‘They were afflicted by a big shock, i.e. they were very scared’)

This analysis provides an interesting contrast to the case presented by Tognini-Bonelli (2001, 115). In her analysis of the semantic prosodies of the Italian lemma *andare*

incontro she indicates that the semantic ambiguity of this lemma is solved on the colligational level: the literal meaning of ‘going towards someone to meet them’ is always associated with personal pronouns, whereas the metaphorical meaning is characterized by the presence of a particular preposition. In the case of *-swerwe* no such colligational pattern can be identified. The disambiguating factor seems to be vested in the semantic features of the agent itself, rather than in the grammatical patterns associated with the lexical item.

Following the same procedure as was done for *-kgotlelela* (‘tolerate’), the context surrounding the verb stem *-swerwe* is investigated for possible instances of collocation, semantic preference, colligation and semantic prosody. Perusal of the left context does not provide any evidence of any of these patterns except for the presence of a variable patient NP, which also functions as the grammatical subject, as is evidenced by presence of a subject concord. To the right of the verb stem *-swerwe*, the agentive prefix *ke-* (‘by’) appears in all cases, a clear indication that it forms an inherent component of a bigger lexical item and represents a case of colligation, i.e. a grammatical choice. When looking at the collocates appearing to the right of the phrase *swerwe ke* it is clear that this position is dominated by inanimate and mostly abstract nouns. Compare the table below in which the nominal collocates are listed, arranged according to their collocational significance, as calculated by T-score:

Table 1: Collocates of *-swerwe ke*

Noun	Meaning	T-score
<i>tlala</i>	hunger	11.348
<i>boroko</i>	sleepiness	5.374
<i>bolwetši</i>	illness, disease	4.985
<i>lenyora</i>	thirst	3.738
<i>AIDS</i>	AIDS	3.454
<i>phefo</i>	cold	2.986
<i>mpshikela</i>	cold (ailment)	2.828
<i>sehuba</i>	cough, cold	2.643
<i>hlogo</i>	headache	2.537

The semantic preference represented by these items is clear: all of them refer to unpleasant sensations which are experienced on a physical, bodily level. Although not collocationally significant, i.e. with a T-score higher than 2, other nouns appearing in this slot include *papalase* ('hangover'), *segateledi* ('nightmare'), *sehlabi* ('stabbing pain'), *mathata* ('problems'), ~~*segateledi* ('nightmare')~~ and *poifo* ('fear'), thus confirming the semantic preference. For further examples, compare the excerpt from the concordance lines for *-swerwe ke* which appear as appendix B. We therefore have a compound lexical item, consisting of a patient + *swerwe* + *ke* + unpleasant bodily sensation, which represents one single lexical choice. The choice of the verb stem *-swerwe* proscribes a number of other choices, some of which are grammatical, some of which are semantic.

In contrast to the case presented above for *-kgotlelela*, perusal of the context to the left and right of the lexical item does not reveal any specific semantic prosody. However, what is significant in the case of *-swerwe*, is the fact that it is a passive verb, used in a passive structure. The passive in Northern Sotho is widely recognized as a discourse pragmatic structure by means of which the patient is presented as the main topic of the sentence in the preverbal position. Therefore, in passive structures the focus is sharply on the patient, at whom the action expressed by the verb is directed. In this particular instance, the idea of focusing on the patient is so strong, that the so-called active counterparts of sentences such as those in (8) to (10) above are – although grammatically correct – unacceptable from a discourse pragmatic point of view. A sentence such as *Bolwetši bjo šoro kudu bo be bo swere morwedi wa monna yoo* ('A very serious disease has taken hold of that man's daughter') where the patient is relegated to the postverbal position, is unacceptable as active equivalent for the example in (8) above. If semantic prosody is defined as the attitude of the speaker towards a pragmatic situation, then surely the almost abnormal focus placed on the patient in these sentences points towards an element of semantic prosody. It could furthermore be argued that the use of the passive structure not only foregrounds the patient as such, but also implies that the patient is the affected party, afflicted by something that is beyond his control; the victim of outside forces.

This would therefore bring us to the semantic prosody of the lexical item: a (mostly human, or at least animate) subject / patient is through no fault of his afflicted by an

unpleasant bodily sensation. As stated by Sinclair (2004, 34), it is exactly the initial choice of semantic prosody which links meaning to purpose, and all subsequent choices within the lexical item relate back to the prosody. What seems to distinguish this particular lexical item from the previous example and also from the cases presented by Tognini-Bonelli (2001, 106 et seq.) and Sinclair (2004, 30 et seq.) is the fact that the semantic prosody is vested in a particular syntactic structure, i.e. the passive. This should however not be surprising when it is taken into consideration that the passive in Northern Sotho represents a discourse pragmatic strategy, utilized to fulfil a very specific discourse function.

Conclusion

The theoretical implications of the postulated existence of extended, functional units of meaning have to a large extent been covered in the first two paragraphs of this article. Suffice it to say that it has implication for the way in which linguists view the relationship between lexis and grammar / syntax, which are traditionally regarded as related, but separate entities. We have seen that there is a strict interconnection between an item and its environment, which becomes visible in a concordance. As the boundaries between an item and its environment get blurred, it suggests to the linguist the existence of an extended unit of meaning, in which the lexical, grammatical, semantic and pragmatic levels are brought together. Taken one step further, Francis (1993) suggests that syntax is driven by lexis, that lexis is communicatively prior. This link between syntax and lexis has not been fully explored in the current literature and the fact that it is supported by the findings explicated above with regard to the connection between the passive structure in Northern Sotho and semantic prosody represents a significant contribution to the theoretical underpinnings of the notion of a larger unit of meaning. Such an approach furthermore implies, as pointed out by Francis (1993, 143), that syntactic structures and lexical items (or strings of lexical items) are co-selected, and that it is impossible to study the one independently of the other – a fact that linguists need to recognize and more importantly, make provision for in their compilation of grammars. A full exploration and exploitation of the theoretical implications of this approach would very likely culminate in the compilation of corpus grammars, in which the point of departure is lexis, and not grammar.

Secondly, the existence of larger units of meaning would need to be accounted for in language teaching. Teaching of Northern Sotho has in the past been characterized by a grammar-based approach. It was assumed that grammar is the basis of language and that mastery of the grammatical systems is a prerequisite for effective communication. Recognizing the fact that grammar and lexis are not two separate conceptual entities and that the one cannot be studied or taught without due attention to the other, supports the idea put forward by Lewis (1993) that one of the central organizing principles in the teaching of language should be lexis. This in turn would require an overhaul of the methodology of teaching Northern Sotho, and possibly other African languages as well.

In the third instance, when compiling dictionaries, Northern Sotho lexicographers, and by implication lexicographers of the other South African Bantu languages, need to take cognisance of the concept of extended units of meaning. The major structural categories upon which these units are based, i.e. collocation, colligation, semantic preference and semantic prosody can and should be incorporated in the treatment of relevant lemmas. These could be accounted for in an indirect manner by the careful selection of corpus-based usage examples, or they can be addressed directly by means of usage notes.

Corpus-driven linguistic investigation does not yet enjoy the attention it deserves from Bantuists. This study has hopefully paved the way for further scholarly investigation into these languages.

References

- De Schryver, G.-M. 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English*. Cape Town: Oxford University Press.
- Francis, G. 1991. Nominal group heads and clause structure. *Word* 42:145 – 156.
- Francis, G. 1993. A corpus-driven approach to grammar. In *Text and Technology*, ed. M. Baker, G. Francis and E. Tognini-Bonelli. Amsterdam: John Benjamins Publishing Company.
- Hanks, P. 2006. Metaphoricity is gradable. In *Corpus-based approaches to metaphor*

- and metonymy*, ed. A. Stefanowitsch and S. Th. Gries. Berlin: Mouton de Gruyter.
- Lewis, M. 1993. *The lexical approach: The state of ELT and the way forward*. England: Language Teaching Publications.
- Louw, B. 2000. Contextual prosodic theory: Bringing semantic prosodies to life. In *Words in context*, ed. C Heffer and H Saunston. Discourse Analysis Monograph 18. Birmingham: University of Birmingham.
- Nelson, M. 2000. *A corpus-based study of the lexis of Business English and Business English teaching material*. Unpublished PhD thesis, University of Manchester.
- Sinclair, J.M. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J.M. 1996. The search for units of meaning. *Textus* 9(1): 75 – 106.
- Sinclair, J.M. 1998. The lexical item. In *Contrastive lexical semantics*. Current issues in Linguistic Theory 171, ed. E. Weigand 1 - 24. Amsterdam: John Benjamins Publishing Company.
- Sinclair, J.M. 2000. Lexical grammar. *Naujoji Metodologija* 24: 191 – 203.
- Sinclair, J. M. 2004. *Trust the text: Language, corpus and discourse*. Oxon: Routledge.
- Smadja, F. 1993. Retrieving collocations from Text: Xtract. *Computational linguistics* 19(1): 143 – 177
- Stefanowitsch, A. and S. Th. Gries. 2006. *Corpus-based approaches to metaphor and metonymy*. Berlin: Mouton de Gruyter.
- Tognini-Bonelli, E. 2001. Corpus linguistics at work. *Studies in Corpus Linguistics* 6. Amsterdam: John Benjamins Publishing Company.

Appendix A: KWIC lines for *-kgotlelela*

Appendix B: Excerpt from KWIC lines for *-swerwe ke*

ⁱ Northern Sotho is a South-eastern Bantu language (S32) with 4.5 million speakers, the majority residing in the Gauteng and Limpopo provinces of South Africa.

Appendices

Appendix A: KWIC lines for *-kgotlelela*

N Concordance

1 mpša ya bona ... Ke ka lebaka la eng re swanetše go kgotlelela baagišane ba mohuta wo? Naledi Maru Gopola
2 ." 22 Kgoši ga e tenwe ke madume bjalo ka ngaka ge e kgotlelela banyakathušo. Matsatšing a ge Sefenyeko ka
3 ge o phela le batho o tlamegile go ba le pelo e tee ya go kgotlelela batho ba bangwe go etša le ge wena ba go
4 monna ke nku o llaleteng. Ga se thaka ya mogemane go kgotlelela bohloko. Moruti le yena taba ya dipasa o be a
5 le ge a itsebela Modimo. OA e sa le tumo e botse ge le ka kgotlelela bothata le otlwa ge le sentše? Fela, ge le
6 a na le dikutu tše koto tše di tiilego gore di kgone go kgotlelela botšididi bjo olo le diphefo tšeo di tšutlago ka
7 wa dijo o seke wa go feroša dibete. Ka tšwaalo mmele o kgotlelela dijo tša go tonya bokaone. • Dirra gore moo o
8 ka go botša Faro gore o fihlile mo a ka se h'wego a kgotlelela dipolelo tša gagwe. Faro o ile a tabogela godimo
9 mo go ba moloko wa Isiraele; naga ga e sa kgona go kgotlelela dipolelo tša gagwe ka moka. "Gobane Amosi o
10 e lego Maribe James ka ngwaga wa 1912. Morago ga go kgotlelela ditaišego polaseng ya Kwaga, setšhaba sa
11 C Ditshehla di a otšela 69 Maaka: A Bathobaso ba ile ba kgotlelela ditaišo ka bagateledi ba bona. B Bathobaso ba
12 ge di nyakurelwa pele ga Tutu) 69 A (Bathobaso ba ile ba kgotlelela ditaišo ka baetapele ba bona) 70 A (pušo ya
13 ge le ka kgotlelela bothata le otlwa ge le sentše? Fela, ge le kgotlelela ditshokofatšo le le ba go dira botse, gona e ba
14 , go hlompaha thoto ya sekolo lego hlompaha gammogo lego kgotlelela ditokelo tša ba bangwe tša motheo. Mape, go ile
15 tšego, wo o ka kgonago go hloga marega; eupša wa kgotlelela gape le bothitho le komelelo ya selemo. Go na le
16 o homotše ka boomo. Senwamadi o ile a palelwa ke go ka kgotlelela kgatelelo ya monagano wa gagwe. O ile a gerula
17 ya gore bokgoni ke eng, setšhaba se ka se sa kgona go kgotlelela kgethologano yegahlanonng le baithuti yeo e
18 tokollo gore ba fihlele tšego e lego kaone. Ba bangwe ba kgotlelela kgobošo le dikoto, le ditfemo, le kgolego. Ba
19 wo o kotofetšego wa karlone. 18. Crasula - semela sa go kgotlelela komelelo sa matlakala a makoto. 19. Lithops -
20 ke borapolase gore ba hlagise dibjalo tšeo di kgonago go kgotlelela komelelo goba malwetsi. Gore re the re kgone go
21 mo boya bjo botolele matlakaleng bo ka thušago dimela go kgotlelela komelelo. ##### 3. DIPHOOFOLO LE DIMELA
22 sa matlakala a makoto. 19. Lithops - semela sa go kgotlelela komelelo sa matlakala a khupedišwego ka motu
23 e telele ya komelelo, re hwetša mehlaire yeo e kgonago go kgotlelela komelelo, bjalo ka mehlaire ya moyo (baobab) ya
24 dimelameetse le meoka. Dimela tše di bitšwa dimela tša go kgotlelela komelelo. Kamantšho Ntsha e kgolo ye o
25 moše wa mawatlle. Dikgomo tša Afrika - Borwo di kgona go kgotlelela komelelo le malwetsi. Dikeletšo go Balemi 1.
26 ale a pele Phokabošego le ba lapa la gagwe ba ile ba no kgotlelela le ona ao madimabe. Go phela gona ba be ba
27 kgale e be e ile ya fihleletšwa diphelelang gore e kgone go kgotlelela leratadima la fao le bogale e le ruri. Dipente tše
28 mo kgahla kudu. A rata ge nkabe Mahlako a ekwa yena, a kgotlelela lesogana le. O lekile ka mekgwa ye mentšhi go
29 , pele ga badudi ba naga. Setšhaba se ka se sa kgona go kgotlelela mabaka ao go wona katologano ya bokgoni yeo
30 ba tlo boela dinonwaneng. Ge e le wena, phafogela tšohle; kgotlelela mabothata, o dire modiro wa segoledi, o be
31 ke la dikgomo Kejoko magareng ga monna le mosadi, Le kgotlelela madimo le diphefo Le kgotlelela meboto le meedi
32 se dire mabapi le thoto ya tatagwe. Eupša mosadi ga se a kgotlelela maroga ao a go a hwetša, a etšwa go ngwana wa
33 ke malopo. Motho wa batho ruri O ile a sitwa ke go kgotlelela maroga ba a bego ba mo roga ka ona. Ka ge
34 bja nama, le nna nke ke ithctc. ' Gobane le ba go rata go kgotlelela mašilo, mola le le ba bohlae. O'heng le kgotlelela
35 , Le bofile pelo sa pušana ya matongwatongwane, Le kgotlelela mathata le mararankodi a bophelo. 4. 5.
36 le telelele bjo maalta, bjo mo kwagalago le go kgona go kgotlelela mathata mengwageng ye mentšhi. Mo tiragatšo ya
37 yeo e tilego go Modimo le yona e mo kgontšha go kgotlelela mathata ohle. Ke ka tumelo ge mafelalong
38 seema seo se nepišago thalošo ya: 'Sa motho ke go fo kgotlelela mathata. A Maebanamabedi ga a rakwe B La
39 , go ba le tirišano, go ba sethaka gona mešomong ka moka, go kgotlelela mathata a bogadi, bjalobulo sedirišwa sa balaong
40 yeo e tilego go Modimo le yona e mo kgontšha go kgotlelela mathata ohle. Ke ka tumelo ge mafelalong
41 ka tsela ye ngwue, e thata ga O bone ge le kgona go kgotlelela mathata a go swana le a bomakgolo. O re rena
42 . Ke ka fao o hwetšago batho ba bantši ba sa ba kgona go kgotlelela mathata. Monna yola wa batho o itše ka ota le
43 ga monna le mosadi, Le kgotlelela madimo le diphefo Le kgotlelela meboto le meedi ya bophelo Go tšaga tšatšing
44 kgontha ya taba ye ka go laetša ka moo Mologadi a ilego a kgotlelela meleko ya Lekope. (10) (10) [40] GOBA 3.2 Bala
45 gagwe ka mokgwa woo. 'Wa lehlogonolo ke monna wa go kgotlelela melekong; gobane ge a bonwe go botega, o tlo
46 ba be ba tseba gore motšweletši wa bona ga a ke a kgotlelela melepologo efe, ga ge kgathalege gore motho ke
47 gore a le a mo thuše wona. Sentshwiniane a sitwa ke go kgotlelela mengunanguno ya mmagwe, a leka go pitlala
48 ka letswela) Ga se selo tše ka moka! Na ke swanetše go kgotlelela meretiwa ya bophelo ka gobane ke belegwe ke
49 modirong. Motšhišhi, tšhiwana e sa hwego e leta monono. Kgotlelela mohlako wo lehonno, gosasa o tla iteboga.
50 nngwe go lebelela fdse go be go tloša bodutu, ge o be o ka kgotlelela monko wa oli. Ka nako ye nngwe re be re
51 go robala ka gae, mo mosadi yo mongwe a kilego a re go kgotlelela monna wa go se je ka gae mantšiboa a no fela a
52 a šiša. Gopolang mehla yela ya ge le sa tšo bonegetlwa, la kgotlelela ntwa e kgolo ya ditaišego. Mo gongwe ba le
53 wona, gone fao a hwetša go se bokaone bjo bo phalago go kgotlelela phišo ya ka mole letšabeng. E ka ba ka phošo ge
54 se nyakoga ba se ke ba tenega ba tloga. Aowa, ba ile ba kgotlelela pula yeo le ge e be e tonya. 50 Ba ile ba sa letlile,
55 gwa pala mo go ka palago. Mogologolo o re: "Go kaone go kgotlelela segoboga se o se tsebago, e sego se o sa se
56 ka meetseng a mogobe, ba rata go bona yo a kago kgotlelela sekhutamoya go feta ba bangwe. Na ga ke re ba
57 o re bolaila thotho tše! Ge nka be e se yena a go no kgotlelela selo se ba rego ke sekolo nka be gona bjale re
58 yona. Tabakgolo ke seroba nka se kgotlelele seroba. Nka kgotlelela serupa bjang mmele o baba? (A befiša
59 ke go kgotlelela, gomme ngwana wa batho le yena a kgotlelela sesadi. Ge go iwa gae batho bale go bego go
60 ke manyami kudu, o sa tšwa gona bjale. Le nna ga ke sa kgotlelela taba ya monna wa go phela a tšwile." "Ke kwele
61 ke leina le lengwe la Tebogoo? O ra gore Motsomoi o tla kgotlelela taba ye ge a ekwa? Tsentse Mphela a Malope ke
62 a fohla badimo ga a bolawe. Ge re go thongka seso kgotlelela thokwa, Gobane o se ngwale le fiša tšhiřitširi;
63 noga moleteng, ya tšwa ya yo sela. Naa motho yena a ka kgotlelela tšala? Aowa, a ka se e kgotlelele, g(mnie .
64 maabane, ba be ba tlo dio kwa ka yena. O be a ka mpa a kgotlelela tšala le phefo, e sego tšhila. Batšwadi ba gagwe
65 Kgoši o a tseba gore Marutla ke motho wa go se kgone go kgotlelela tšala. Ke bone gore o lapile kudu ka baka la tšala.
66 re fetša. Aowaowa. ' Mosadi wa Segatamoroko ga a ka a kgotlelela tšontlo ya lapa la bona. O be a setše a eme
67 etla! O realo Morena wa mašaba. Gomme ke mang a tlogo kgotlelela tšatši leo la go tla ga gagwe? Ke mang a tlogo

Appendix B: Excerpt from KWIC lines for *swerwe ke*

N Concordance

205 ke bolwetši bja motšatši. Gonabjale bana ba swerwe ke khwaši, dimpana ke mararampana
206 botšiša noga. Phuti e be e lapile kudu, mme e swerwe ke lenyora. Leleme la yona le be le
207 ka moka, gomme ge bošego bo fihla ke ge di swerwe ke tlala. "Nna ke tseba gore re ka
208 tša monyanyeng wa Kokolohute o tla fetša a swerwe ke bolwetši bja pelo. O tshwenyegelang
209 le lengwe mmutla o be o orela bothunya. O be o swerwe ke tlala, eupša o be o tšwafile go nyaka
210 Katse e ile ya nagana ya hwetša gore e swerwe ke tlala. Ya fetola ka go botšiša: "Na o
211 Ke swanetše go lla ka gore mohlomongwe o swerwe ke bomaemae. "Ga ba mo swara. O
212 di šetšego le go bona gore o tla ja eng ge eba o swerwe ke tlala. Yena o be a sa dume selo ge e
213 . A di nape di fele moratiwa, nna ke lapile ke swerwe ke boroko. "A edimola a ikotlolla a boa a
214 a robetše. O letše a pitikologa bjalo ka motho a swerwe ke teng goba a longwa ke ditšhitšhiri
215 ebile ge a šetše a go swere O tla ba wa re O swerwe ke mpša ka dipolai. Ntahle ya gagwe
216 ka gore ke sediba mang sona se o rego ge o swerwe ke lenyora se go gane ka meetse? O ile
217 nyaka bogobe. Ka lora gape a re ke mo fe dijo o swerwe ke tlala. Maloba o rometše Nkokona
218 ga a na lapa. SEKGOLE: Mphe bogobe ke swerwe ke tlala. Bogobe bo kae? MOKGADI:
219 le go fula matlakala a mehlare. Naledi le yena o swerwe ke marena letšatši le ruthetše. O tletše
220 gabedi, gararo, o bile o tennwe. Bontši ba swerwe ke tlala. Bothata ke gore ga go lešobana
221 gantši se se dirwa ke phišo e bile ge motho a swerwe ke lenyora o nwa meetse a mantši lentšu
222 ka dinkong, gantši a ba gona ge motho a swerwe ke sehuba goba mpshikela selo seo go
223 ka gae motho wa go phela a le bodutwana e re o swerwe ke boroko, wa go dira dilo ka go
224 mathoko a mahlo, gantši o laetša gore motho o swerwe ke bolwetši bja mahlo go šutašuta le go
225 teng, gantši ge motho a lapile monagano goba a swerwe ke boroko lentšu leo le šomišwago bjalo
226 hle, monna. Mphobadimo še. MADIMETŠA (o sa swerwe ke boroko): O reng? Ke eng?
227 ya ka. Ba tla itshola ka morago. Mahlo a ka a swerwe ke boroko, fela mogopolo wa ka ga o na
228 šo, ngaka, O laodiša gore ga a lala a robetše, o swerwe ke sehlabi mo tšase ga letswele, gomme
229 šole ka mola, o re ga a lala a robetše, o swerwe ke sehlabi se sengwe seo se mo fetšago
230 la o tsogile bjang) Ke a babja ngwanaka, ke 83 swerwe ke sehlabi se se soro mo (O mo šupetša
231 a sa robetše. META:(O a phakgama, o kwa a swerwe ke tlala) MmagoNgwato! MPUME: Mma!
232 : Mma! META: Mphe sa go ja ngwanaka, ke swerwe ke tlala. Bjale gona ke kwa ke le kaone
233 Lekgala a re o ngadile le dijo. Ge a tloga a re o swerwe ke tlala, ga se ka lefeela, there-to o di
234 92 LEKGALA: Gona moo, mokgekolo ge a swerwe ke bolwetši, a gopola gore ge a be a sa
235 Rebecca o gotša mollo: Go a sa (b) Re a ja: Re swerwe ke tlala (c) Ba tla goroga: Letšatši le a
236 ganelela. Go edimola: Ka molomo. Ge motho a swerwe ke boroko goba a lapile. Go ithimola: Ka
237 goba a lapile. Go ithimola: Ka dinko ge motho a swerwe ke mpshikela. Mogwang: Ke mahlaka ao
238 ka letswai. • Banna ba lema tšhemo. • Bana ba swerwe ke tlala. • Jesaya o bolaya phala. •
239 a mangwe ka nako, o tla swamoga moya. O swerwe ke tšhipisepanšrš sa Makgowa. A
240 ke boro]] Go ithimola: Ka dinko ge motho a swerwe ke mpshikešo goba a lapile. Mogwang:
241 ke bonna, ba baka sebešo ramamilana, ba swerwe ke phefo, mošemane yo mosese, ba
242 o ile a tsena, a iša mahlo. Mong wa ntlo o be a swerwe ke boroko bjo bongwe, fela a sa kwe selo
243 fihla motseng wo mongwe. Ka nako yeo o be a swerwe ke hlogo ya mabjalwa a go
244 batala ba re motho ge a ka tšwa ka gabo a swerwe ke tlala, le kua a yago o tla hwetša tlala,
245 gore mmagwe o mmoditše gore ge a ka kwa a swerwe ke bolwetši bjo bobjalo a tsebe gore
246 lebaka leo ge Mmaphuti a tsena, ke ge tau e swerwe ke segateledi. E itše ge e phafoga ya
247 wa monna yoo. Morwedi lwa monna yoo o be a swerwe ke bolwetši bjo šoro kudu. Pheko ya
248 , gomme ka lebaka leo mogongwe O be a swerwe ke segateletši. Ba ile ba bona gore ke