
Narzędzia korpusowe

Beata Wójtowicz

Anotacja = znakowanie korpusu
vs czysty tekst

Anotacja

Analiza tekstu — morfologiczna, składniowa, wreszcie semantyczna.

Tekst dostępny urządzeniu analizującemu jest obiektem unilateralnym; jego rozbiór to bilateralizacja. Automatyczny analizator uważać można za model odbiorcy tekstu. Wyrażenie językowe zostanie zbilateralizowane (odebrane, zanalizowane, zrozumiane), jeśli jego składniki rozpozna się właściwie →

- lematyzacja

- problem homonimii

Korpus poddany częściowej choćby bilateralizacji nazywa się korpusem znakowanym (ang. *tagged corpus*).

Leksem

- Jednostka opisu słownikowego języka, wyraz słownikowy.
 - Zbiór form wyrazowych przeciwstawiających się sobie pod względem kategorii fleksyjnej.
 - Formy *obraz*, *obrazem*, *obrazy* reprezentują leksem OBRAZ.
 - W przypadku leksemów nieodmiennych są to zbiory jednoelementowe, np. PRZECIW, NATOMIAST.
-

Terminy

- **słowo** — napis ograniczony spacjami lub znakami interpunkcyjnymi;
 - **forma wyrazowa** — słowo z interpretacją gramatyczną i semantyczną (klasa gramatyczna, wartości kategorii fleksyjnych, nazwa jednostki słownikowej);
 - **leksem** — zbiór form wyrazowych o tym samym lub regularnie zróżnicowanym odniesieniu do rzeczywistości (i, w ideale, o podobnym kształcie)
-

Homonimia

- Równokształtność znaków językowych o wyraźnie różnym znaczeniu.
 - Homonimia całych jednostek słownikowych (leksemów), np. *bal* 'zabawa' i *bal* 'kloc'.
 - Homonimia form wyrazowych należących do różnych leksemów.
-

-
1. tonie: Czasownik, 3 osoba, liczba pojedyncza, czas teraźniejszy (TONAĆ)
 2. tonie: Rzeczownik żeński, mianownik, liczba mnoga (TOŃ)
 3. tonie: Rzeczownik żeński, biernik, liczba mnoga (TOŃ)
 4. tonie: Rzeczownik żeński, wołacz, liczba mnoga (TOŃ)
 5. tonie: Rzeczownik mesko-rzeczowy, miejscownik, liczba pojedyncza (TON)
 6. tonie: Rzeczownik mesko-rzeczowy, wołacz, liczba pojedyncza (TON)
 7. tonie: Rzeczownik żeński, celownik, liczba pojedyncza (TONA)
 8. tonie: Rzeczownik żeński, miejscownik, liczba pojedyncza (TONA)
-

Homonimia to równokształtność form wyrazowych należących do dwóch leksemów (jest to homonimia międzyparadygmatyczna w nomenklaturze Elżbiety Awramiuk (1999: 7)).

łapie → łapać/łapa

Homoforma homonimiczna to takie słowo, które jest wykładnikiem tekstowym form wyrazowych należących do dwóch leksemów. Parami form homonimicznych są na przykład pary

(1)-(2), (1)-(5), (1)-(8).

Synkretyzmem nazywa się równokształtność form wyrazowych wewnątrz jednego leksemu (homonimia wewnątrzparadygmatyczna, Awramiuk (1999: 7)).

Homoforma synkretyczna to takie słowo, które jest wykładnikiem tekstowym form wyrazowych należących do jednego leksemu.

Synkretyzm ilustrują np. pary (2)-(3), (5)-(6), (7)-(8).

Dana homoforma może reprezentować:

(a) parę form wyrazowych synkretycznych,

(b) parę form wyrazowych niesynkretycznych (każda o jednej interpretacji),

(c) parę typu „forma wyrazowa niesynkretyczna — forma wyrazowa synkretyczna”.

tonie:

celownik **lub** miejscownik, liczba pojedyncza (TONA)

miejscownik **lub** wołacz, liczba pojedyncza (TON) (ad (a))

gra:

mianownik, liczba pojedyncza (GRA)

3 osoba, liczba pojedyncza, czas teraźniejszy (GRAĆ) (ad (b))

pierze:

3 osoba, liczba pojedyncza, czas teraźniejszy (PRAĆ)

mianownik **lub** biernik, **lub** wołacz, liczba pojedyncza (PIERZE) (ad (c))

Zinterpretowanie homoforny polegające na przypisaniu jej odpowiednich leksemów to **dehomonimizacja (lematyzacja, hasłowanie)**.

Zinterpretowanie homoforny polegające na przypisaniu jej odpowiednich charakterystyk gramatycznych to **desynkretyzacja (dehomonimizacja)**.

Anotacja morfosyntaktyczna

- Poszczególnym ciągom znaków (słowom) przypisane zostały znaczniki interpretujące dane ciągi jako wykładniki tekstowe pewnych form wyrazowych oraz ich charakterystykę morfoskładniową.
-

- Klasa gramatyczna = część mowy
- Kategoria gramatyczna (na przykładzie korpusu IPI PAN)

- Liczba
- Przypadek
- Rodzaj
- Osoba
- Stopień
- Aspekt
- Zanegowanie (pisanie vs niepisanie)
- Akcentowość (jego vs go)
- Poprzyimkowość (niego vs jego, go)
- Akomodacyjność (dwaj vs dwóch)
- Aglutynacyjność (niósł vs niosł-)

Narzędzia komputerowe

- Analizator morfologiczny
 - Amor
 - Morfeusz
 - Analizator składniowy
 - **Świgr** – generuje drzewa składniowe, gramatyka programu wywodzi się z Gramatyki formalnej języka polskiego Marka Świdzińskiego (1992).
 - **Słownik gramatyczny języka polskiego** - zawiera charakterystykę gramatyczną ponad 256 tys. wyrazów (leksemów) polskich.
-

Wyszukiwarki korpusowe

- Programy 3. generacji
 - WordSmith, MonoConc, AntConc, Xaira
- Programy 4. generacji
 - wyszukiwanie przez interfejs internetowy
 - PELCRA

Język zapytań

- Wyrażenia regularne
- Dedykowany

British National Corpus – składnia zapytań

- **cat _ dog** finds three word phrases of which the first word is `cat' and the last is `dog'
 - **!cat dog** finds occurrences of `dog' not preceded by `cat' within the same document
 - **cat*dog** finds occurrences of `cat' followed anywhere within the same document by `dog'
 - **cat#dog** finds occurrences of `cat' followed or preceded by `dog' anywhere within the same document
 - **cat*dog/10** finds occurrences of `cat' followed by `dog' within ten words
 - **cat*dog/** finds occurrences of `cat' followed by `dog' within a single <head> element
-

cat _ dog

[ADF](#) 317 Solitary people often have a **cat or dog** as a companion too; but it is not the same as having a companion of one's own kind.

[ADF](#) 383 The horse stabled in a shed or barn cannot see people, cars, and animals come and go; or a cat or dog wandering by, birds flying, the wind in the grass and trees, and s

[ALL](#) 2110 South towards Cardonald, with the Cat and Dog Home in the distance, lay the facade of the Palace of Industry, from this height, a gentle curve with long glass galleries Bellahouston Drive.

[ANK](#) 651 He, Drew, was stroking a car as if it were a cat or dog.

[AYK](#) 1856 Next time you open a book, stroke a cat or dog or lie between cool sheets allow yourself to become aware of the texture and the temperature of what you touch.

[B03](#) 2045 Each of the Pet Value policies — insuring a cat or dog — including one year's cover for veterinary fees, both conventional treatment and complementary medicines — herbal and homoeopathic medicine — up to £1,875.

[B22](#) 1544 Called 'Suspense', it showed a small girl, saying her prayers before having breakfast in bed, on either side her cat and dog waited in great anticipation.

[BLY](#) 1293 I was particularly struck by the requirement for a preserved cat or dog.

[BMS](#) 262 Bereaved pet owners no longer have to go to a cemetery to give their cat or dog a loving send off.

[BNG](#) 2200 Strong, but not 'too heavy to move', covers must be provided unless the area is absolutely cat, dog and vandal proof.

[BNY](#) 1657 For instance, the ground you are netting may have been worked by a fox, cat or dog just before your arrival on the scene and the rabbits sent scurrying for home.

[CBC](#) 752 I even bought a set of those flash cards with pictures and words such as Cat, Dog, Rabbit, Mummy and Daddy in my attempt to awake in him a hunger for knowledge.

[CES](#) 1479 Whoever it is made a start with the cat and dog, the family pets.'

[CMS](#) 554 Driving instructors will tell us that if a cat or dog strays into our path, and we are unable to stop or otherwise avoid it in an orderly fashion, then we should run it down.

[EF1](#) 1992 We'd fight like cat and dog.

[FAC](#) 394 For instance,cat and dog have a fairly high degree of paradigmatic affinity, as they are equi-normal in a wide range of contexts:

[FAC](#) 1760 animal :cat ,dog ,lion ,elephant ,aardvark , etc.

[G36](#) 2154 Well, are you a cat or dog person?

can=VVI (the infinitive form of lexical verbs)

Results of your search

Your query was

can=VVI

Only 4 solutions found for this query

[BIM](#) 703 We can also return any cans through the 'Save-a-Can' can banks which are becoming increasingly common.

[BNG](#) 1424 Only when children can do this confidently can counting have any meaning, as counting is simply matching objects (or people, or whatever is being c

[ECN](#) 797 Second, the instrumental aspect of order is valued so that people who want to can 'get on'.

[KCP](#) 8810 if we're gonna can I have another line, didn't get the right one here !

Helsinki Corpus of Swahili

- Nyumba
 - Nyumba*
 - *umba
 - m.mba
-

*umba

kukabwa ilimalizwa wakati nahodha wa Mapipa , Juma **Chilumba** , alipoingilia kati na kufanikiwa kuwatenganisha na
za ujumbe wa NEC , bado amelazwa katika **Chumba** cha Wagonjwa Mahututi , Icu mjini hapa .

uhi wakati wenye chumba wamekwenda kuhangaikia maisha . **Chumba** kilichoungua ni namba 26 cha jengo ' N

Akasema kuwa waganga hao ambao kituo chao kinaitwa **Dumba** wanu kliniki senta , ujumbe wao uliongozwa na
hiyo imetolewa jana kwa vyombo vya habati na **Fumba** , na kusisitiza kuwa hatambui kadi mpya za
Kamati ya Michezo Wilaya ya Ilala , Mwamvita **Fumba** . Akizungumza kwa njia_ya_simu leo asubuhi , Katibu
kauli ya Kamati ya Michezo ya Ilala . **Fumba** ambaye pia ni kaimu msajili msadizi , jana
za Simba na hawazitambui . Dewji alisema , **Fumba** amekuwa mvurugaji mkubwa wa amani za klabu ,
9 mwaka huu . Akikaririwa jana Jijini , **Fumba** amesema kuwa , uchaguzi mkuu wa klabu ya
wa Simba umepangwa kufanyika Juni 8 , ambapo **Fumba** amesema kuwa kadi zitakazotumika ni zile za zamani
Msajili wa Vyama Wilaya ya Ilala , Mwamvita **Fumba** kuwa Uchaguzi Mkuu wa klabu ya Simba ,
katika maeneo huru matatu ya Unguja Mjini , **Fumba** na Pemba Micheweni , " akasema . Akasema
zote za serikali hutibu wanachama wa Bima - **Humba** . Na Seraph Kuandika , Morogoro . Mfuko
Kati . Duka la nguo linalomilikiwa na Bwana **Humba** Hashim lililoko mtaa wa Kongo nyumba namba 111

m.mba

kana na tatizo hilo . Tusaidie jamani , **mamba** watatumaliza , wananchi walio . Na Saic
maji ya mto Zigi ambao umekuwa na **mamba** wengi ambao walikuwa wakitafuna watu
uruwa amesema mto huo umekuwa na **mamba** wengi hali ambayo imekuwa ikiwafanya
na kuona picha za watangazaji au za **mamba** yanayotangazwa . Kuratibu : kufanya kitu
shahidi . Ikiwa dini na imani inapenya **mamba** yote katika maisha yao , bila_shaka wata
li zilizomo humo kama_vile samaki na **mamba** ziwe sehemu mojawapo muhimu za uch
ariti mengi . Hata_hivyo , nchi hiyo ni **mamba** wa NATO uanachama ambao wachambi
gewahusisha " Waislamu " **mamba** wa baa hiyo . Kwa_mujibu_wa waandaz
zake waliendesha maisha yao wakiwa **mamba** wa jamii hiyo . " Tumuangalie Mtu
walikamatwa na kuwekwa kizuizini na **mamba** wa vikosi vya ulinzi na usalama na tume
walikamatwa na kuwekwa kizuizini na **mamba** wa vikosi vya ulinzi na usalama na tume
sababu inayomfanya achelewe kupata **mi mba** # "\$ Baada_ya maelezo hayo ya m
eze kutunga sheria inayoruhusu utoaji **mi mba** " , alisema Bibi Mgaya . Habari za
usababisha maambukizo katika mji wa **mi mba** " , alisema . Aliongeza kuwa , utoaj
wanapandikiza fikra za " uzuiaji **mi mba** " - ambazo hazilingani na uwajibik
isha ya nguruwe ; na wote wawili tuna **mi mba** " . " Mama yangu wee # "
yanaangamizwa na wazazi wanaotoa **mi mba** " . Ndugu zangu Wakristo inasikitish
mumewe ili waelewane wataifanyaje **mi mba** " yao " hiyo . Sambamba na l
' . ' Nimewacha mke wangu na **mi mba** ' Mbinu hizi za majibizano zinasaidia sar
o Ulaya na Marekani vituo vya kutolea **mi mba** (Abortion clinics) vilianzishwa hasa katik
auri na madawa kwa wanaotaka kutoa **mi mba** (abortion) ili wajipatie pesa bila kujali
moja_na damu ya mwezi yaani kutoa **mi mba** (abortion .) Taasisi ya utafiti juu_ya
balimbali ili kuonesha kitangulizi cha **mi mba** (embryo) cha mwanadamu kwanza kilio

Składnia zapytań - anotacja

- [bf="sema"]
 - [bf='ua' pos='v']
 - [bf='ua' pos!='v']
 - [bf='ua' pos='n']
 - [bf='ua' msd='1/2*']
 - [pos='v' msd='*caus*']
-

[bf=“sema”]

Awali kwa_sababu haina nguvu zozote kisheria . **Akasema** pia marekebisho aliyofanya mlalamikaji katika maombi
nuuguzi kuvaa kitambulisho kifuani linaendelea . **Akasema** rushwa imeshamiri zaidi katika kitengo cha uzazi ambao
sekta binafsi , elimu , afya na_kadhalika . **Akasema** suala lingine linalokwamisha wawekezaji wa makampun
na watoto wasitozwe ada kwa_ajili_ya matibabu . **Akasema** vitengo vingine ambako rushwa imeshamiri ni chumba
fanye " mlalamikaji alimnukuu mshitakiwa . **Akasema** wakati huo alikuwa amebaki na upande mmoja wa
mbilia sehemu za miinuko kwa kuogopa mafuriko . **Akasema** walifunga vyumba ama nyumba zao ili kuzuia vitu
vbazi mkoani kwake yanaongezeka siku hadi siku . **Akasema** yote hiyo kwa_sababu mkoa wake upo katikati_ya mikoa
milioni kumi katika wilaya zake zote tano . **Alisema** , juzi peke_yake ukanda wa Kijani pekee ulipandwa
no huo ambao huenda ukawang'oa madarakani . **Alisema** , kwa_mujibu_wa katiba ya Yanga ya mwaka 1992
ayoanza mwezi Oktoba na kumalizika Desemba . **Alisema** Oktoba Mosi , bingwa wa taifa wa uzito
wa Yanga na kuwashawishi kufanya mgomo huo . **Alisema** Rage amechanganyikiwa baada_ya watu kugundua kuwa
majina ya wachezaji 25 na viongozi watano . **Alisema** TBL ilitoa sh. milioni 7.5 , lakini cha
miaka ya zamani " , alisema Ramadhani . **Alisema** Taaa iliwahi kuwaomba wabunge kuchangia sh. 2,000 kw
ri_karibuni , tena mapema iwezekanavyo " . **Alisema** Yanga haiwezi kuacha kufanya mazungumzo na Kibo kw
yake ya baadaye ataanzisha kampuni ya uwakili . **Alisema** akishaachana na masuala ya usajili wa vyama vya
(BMT) ili lishughulikie suala hilo . **Alisema** hadi sasa tatizo kubwa ni vifaa vya mazoezi
atoe msamaha ili acheze mechi hiyo " . **Alisema** hadi sasa wachezaji wote , akiwemo Mhando ,
akisetiriwa na wazazi wake (mdai) . **Alisema** hatua za awali alizochukua kabla ya kufikisha suala hilo

[bf='ua' pos='v']

h maiti ikamsukuma kumwua nduguye . **Akaniwua** akawa imongoni_mwa wenye nasara : hapa mwenyezi k
alikuwa ni Bw. Agathon Simbaulanga . **Amuua** mkewe leo alfajiri . Na Keregero Keregero ,
. Uhai wa mtu ni kitu kitakatifu . **Anayeua** mtu mmoja pasi na haki ni sawa na
ra kuishambulia Ireland ya Kaskazini ? **Ikawaua** wa-irish walio Wakatoliki na Waprotestanti ? Marekani ina
engine watatu . Ajali Ya Gari La Harusi **Iliyaua** Watano : Wakazi Tanga roho juu kwa kudhani
ifu katika ardhi , inadai inatengeneza . **Inaua** na kubomoa nyumba za Waafghan kwa mkono wa
za mitaa . Kesi Ya Daktari Anayedaiwa **Kuuu** Mgonjwa Nesi ababaika ... atoa ushahidi unaojikanganya Na
i Yetu . Tusipumbazwe huu ni unafiki . **Kuuu** ni kuua . Uwe umeua Jumamosi au Ijumaa
ionesha wazi kulihami Kanisa . " **Kuuu** yaelekea ilionekana itakuwa komesha asingethubutu tena mtu
. Pindupindu larudi Dar kwa kishindo . **Laua** watatu ... wengine hoi wodini . Jumamosi ,
igizwa katika Qur-an : Enyi mlioamini ! **Msiue** mawindo na hali ya_kuwa mmo katika Hija au
ujumbe huo , Issa ameandika : " **Najiua** kwa uamuzi wangu , sikushawishiwa na mtu yeyote
gine haikukubaliwa . Akasema : " **Nitakuua** " . Mwenziwe akasema : " Mwenyezi huwapoke
a . " Kijana yule alisema : " **Nitakuua** " kuwa ana uwezo wa kufanya hilo !
a mjini hapa Alhamisi ya Februari 27 . **Nitamua** Palasa - Matumla . Na Majuto Omary ,
waka huu . Meneja YANGA Ajigamba ' **Tutsua** sana mnyama mwaka huu ' Na Robert Komba

[bf='ua' pos!='v']

_na kodi zinazolipwa na makampuni hayo . Dk. **Maua** aliyataja makampuni hayo kuwa ni pamoja_na Mobitel ,
37) , mkazi wa Kijitonyama kwa Ali **Maua** ambaye ni muuzaji wa duka hilo la Karishina
baba wa kumtunza , mama yake mzazi , **Maua** amefungwa jela miezi sita kwa kumpiga mwanamke mwenz
sema baada_ya kumaliza kutumbuiza Bukoba , **Maua** atarejea Dar_es_Salaam ili kumaliza albam yake mpya itak
Naibu Waziri wa Mawasiliano na Uchukuzi Dk. **Maua** daftari ambao walishirikiana bega kwa bega na madaktari
izi chini_ya 30 shule inayoongoza ni Seminari **Maua** ikifuatiwa na Seminari Sengerema na Seminari ya Mtakatifu
wanafunzi wake 19 , hakuna . Seminari ya **Maua** imeshika nafasi ya 9 na mwaka uliopita ilikuwa
za seminari , Dk. Nkumbi alisema Seminari ya **Maua** inaongoza katika kumi bora kundi hilo . Alizitaja
nyumba vinne . Ilikuwa imejengwa madhubuti . **Maua** kadhaa ya kupendeza macho yalizunguka pembeni . Pia
ewe kutosikia kilio chake ndicho kilichopelekea **Maua** kubugia midonge hiyo ya uchungu ambayo idadi yake
shwa na dereva Mohamed Hussein kutokea Ali **Maua** kuelekea Sayansi . Kamanda amemtaja aliyefariki kuwa ni
hafila hiyo . Monday , November 04,2002 . **Maua** kukaribisha Mwezi Mtukufu kesho . Na Mwandishi Wetu
ulia tukio hilo , Kamanda akasema chanzo cha **Maua** kukatiza uhai wake ni wivu ambapo alikuwa akimtuhumu
16,2002 . Miss Mashariki , Kinondoni leo . **Maua** kuzindua Chenkula leo . Na Ramadhan Libenanga ,
16,2002 . Miss Mashariki , Kinondoni leo . **Maua** kuzindua Chenkula leo . Na Ramadhan Libenanga ,

[bf='ua' pos='n']

yahali palipofanana na bustani . **Maua** ya bustani hii yalikuwa meupe na yalitoa mwanga hindu huabudu mti wa Ashoka ? **Maua** ya mti huu huchumwa , kisha maua hayo chafya . Waliutazama msalaba . **Maua** yaliwadondoka . Niliyakumbuka maneno ya Biblia kirika alikuwa nje katika bustani . **Maua** yaliyokuwa yamechanua yalivuta macho na harufu che . [1] Kama miche ya waridi , **Maua** zabarajudi , Matawini kocho koche . Mjinga \$ imu wao kivitendo . Habari za Bi **Maua** zilifikishwa kwa mara ya kwanza An-nuur na mama wa kwa Umoja mpya wa Afrika (**Ua**) na kuitaja kuwa ni hatua ya kihistoria i kamisheni ya Umoja wa Afrika - **Ua** , Amara Essy ametetea haja ya kuongezwa misaada yote ili nafsi yako itakatike . b - **Ua** , langu , sema kauli ya mwisho ili a kutiwa njiani Umoja wa Afrika - **Ua** . Mzozo wa Madagascar unatishia kuugubika mkutano i_ya kivuli cha Umoja wa Afrika - **Ua** . Yote yanakubaliana maendeleo yameweza kufikiwa : Udongo mwekundu . Kitunda : **Ua** ... Watu : Udongo mwekundu . (Watu : Udongo mwekundu . Kitunda : **Ua** ... Watu : Udongo mwekundu . Kitunda : ku za zamani Wakristo walitumia **Ua** aina ya Lily - yungi , kama ishara umuiya mpya ya nchi za Africa , **Ua** kwa wito wa kuitaka jumuiya hiyo kujiingiza kati dini halihitaji uangalifu wa mtu . **Ua** la bustani hutaka sana kutunzwa kama wanawake wa wa katika lugha nyingi , kwa jina **Ua** la jangwani , Desert Flower (The Extraordinary ndelea malkia alieleza . " **Ua** la porini bendine hupita ua la bustani kwa

[bf='ua' msd='1/2*']

Salaam . 30 Desemba , 2002 . Mwalimu [ajiua](#) Dar na kuacha usia wa mali yake .

chama . 06 Januari , 2003 . Mlinzi [ajiua](#) kukwepa mateso . Na Mwandishi Wetu . Mlinzi

Januari , 2003 . . Mfanyakazi wa shambani [ajiua](#) kwa sumu . Na Mwandishi Wetu . Wafanyakazi

kijiji cha Misake , Ihanja , Singida vijijini [alijiua](#) kwa kunywa dawa nyingi ya miti shamba (

toka_kwa ndugu zake zinaeleza kwamba Bosco [alijiua](#) mwishoni mwa wiki iliyopita kwa_sababu zisizojulikana . Ina

aliieleza mahakama awali kwamba mshitakiwa [ali muua](#) Michael Hoya Juni 23,2001 katika eneo la Kichangani

ano ya barabara za Stesheni na Mwanza hoteli [ali muua](#) nduguye Revocatus kwa kumpiga bastola . Kamanda wa

juu ya Tanzania imemwachia huru mwanaume [aliyemuua](#) mgoni wake aliyemfumania ndani_ya nyumba yake aki

rani , huku nikiliazimika kumuona kila_siku mtu [aliyewaua](#) wazazi wangu ? "\$; Serikali ya nchi

shule ya msingi Mkumungu , wilayani Chunya , [amejiua](#) kwa kujinyonga na kuacha ujumbe wenye wosia juu_ya

TTCL) , Julius Mhando (40) [amejiua](#) kwa kujipiga risasi kwenye koromeo . Kamanda alisema

shamba , Aloyce Mselela (30) , [amejiua](#) kwa kunywa kitu kinachodhaniwa sumu baada_ya kuchuki

anchenche wilayani Sengerema mkoani hapa , [amemuua](#) mwanae kwa kumpiga na fimbo kichwani . Mzazi

, lakini benki hii ilikubali . Kuiuza si [tunawaua](#) wakulima ambao hii ndiyo tegemeo lao ? "\$

2000 saa mbili usiku , washitakiwa kwa pamoja [wali muua](#) Bariki Metisho kwa kumkatakata kwa mapanga . Kesi

ko eneo la Bugando maghorofani watuhumiwa [wali muua](#) Hassan Yusuf . Aliendelea kudai kuwa katika shitaka

sa kudai kuwa katika shitaka la pili watuhumiwa [wali muua](#) Patrick Terespory na katika shitaka la tatu watuhumiwa

natatizo mbali_mbali . Utasikia vijana wengine [wanajiua](#) kwa_sababu_ya kusemwa kidogo tu na wazazi wao wana

apadre na kuwalazimisha watoe fedha la_sivyo [wanguwaua](#) . 20 Januari , 2003 . Aomba wakimbizi

ike , Chenge anasema kuwa wagombea binafsi [wataua](#) upinzani kwa vile haujakomaa . Kwa_hiyo anataka wapinz

[pos='v' msd='*caus*']

za Mitaa , Dk. Walid Amaan Kabourou . **Kuthibitishwa** kwa Uenyekiti wa Dk. Kabourou kunafanya kwa mar
oja wa wananchi waliofika kusikiliza mjadala huo . **Kuwasilishwa** kwa mjadala huo kwenye baraza kunafuatia kauli y

12 kutoka kila pembe ya Afrika . " **Ni mefurahishwa** sana na uwakilishi wako mzuri katika mashindan
sha kuwasiliana na watafiti wenzake duniani kote . **Ni meingiza** shughuli zangu kwenye kompyuta tangu mwishoni mv

Hamidu , Thabit Bushako , Sanifu Lazaro ' **Tingisha** ' , Abdallah Msamba , George Lucas ,

katika soko hilo , amedai na kuongeza . **Tumechoshwa** na tabia ya watendaji wa Tume ya Jiji

Afrika , ITV , Dtv , CTN . **Wakatisha** maisha yao kwa kitanzi cha katani Mwingine agongwa

kocha wa klabu ya Majimaji , Nzoyisaba Tazany **aifundishe** timu ya soka ya Simba . Mmoja wa

ibu kuwashawishi wamchukue kocha huyo Moongo **ai marishe** benchi la ufundi la Simba . Kwa sasa

amesomewa shitaka na W/p Hadija Msangi , lakini **akakanusha** shitaka hilo . W/p Hadija ameambia mahakama kuw

Siku hiyo nilikuwa katika siku zangu ... mchungaji **akanilazimisha** na kuniingilia ... " anasema binti huyo katik

. Baada_ya kukata hukumu hiyo , hakim pia **akatahadharisha** kuwa iwapo ataendelea na hali yake ya kupand
imisha taifa dhidi_ya ukimwi " . Bw. Mazzuki **akawachekesha** wanasemina pale alipofananisha mpambano d

mikoani , amekuwa mlevi kupindukia na amekuwa **akibadilisha** wanawake kama khanga . Akaendelea kudai kuwa ,

aina ya Honda Znz 4713 cc 125 aliyokuwa **akiendesha** akitokea Mbeya mjini kwenda Tukuyu kugongana usc

Internet

- Programy konkordancyjne
- Teksty – *web as a corpus*

Konkordancja na tekście, który wkleimy

<http://ec.hku.hk/vocabulary/concordancer.htm>

Concordancer

This program searches words from a text that you paste into the word. Enter the text in the big box below and search word in the small box. You can use OR (in capital letters) for boolean searches.

We shall discuss what a corpus is by looking at how the word is used, in particular by linguists. What kind of an object is a corpus, and what is it likely to be useful for?

We learn the sense of a newly-encountered word in different ways. Young children experimentally combine or mutate words to see which uses meet with approval; older ones do the same in the process of defining peer groups based on a shared exotic vocabulary. In

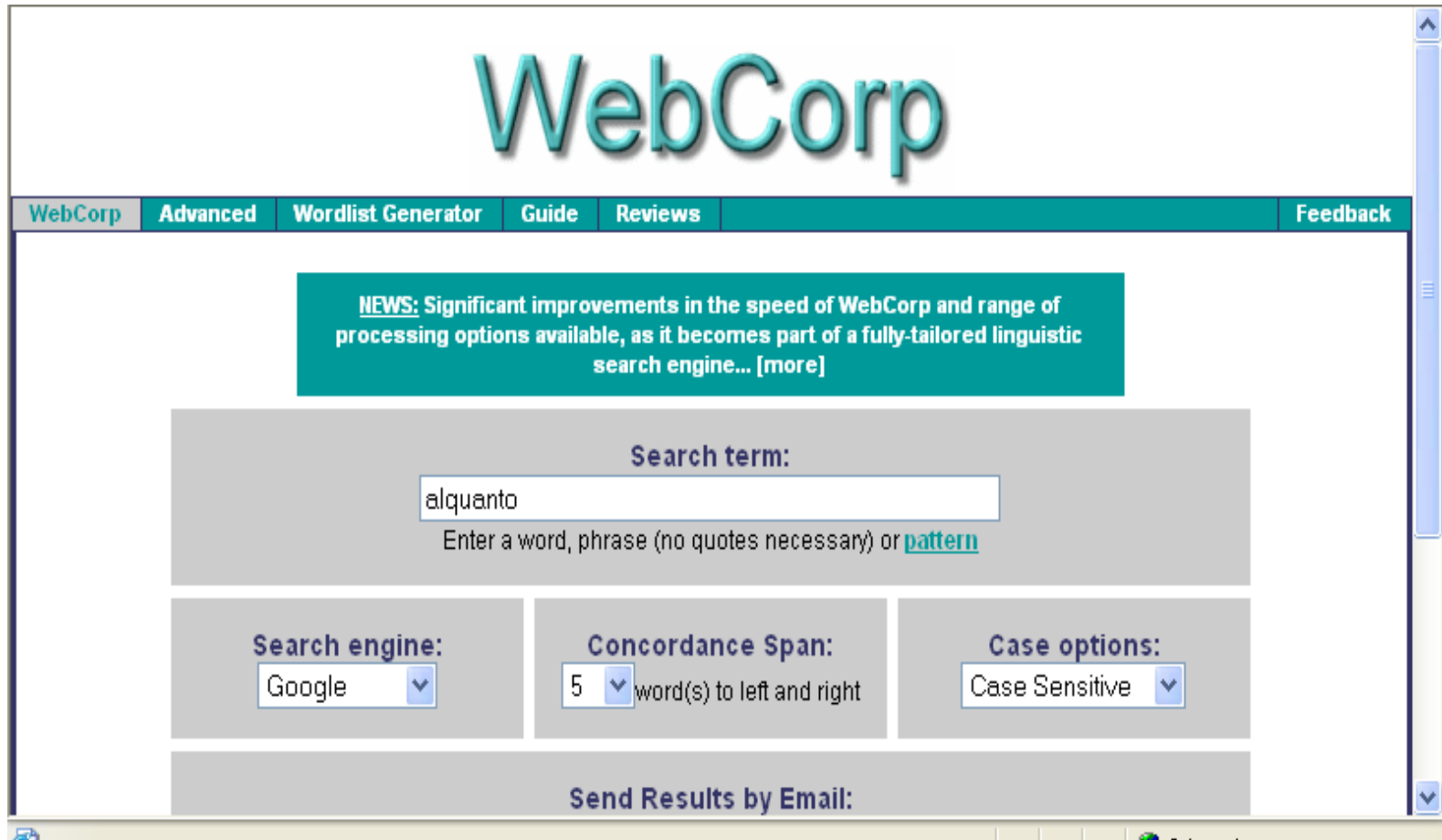
word:

location:

sort by:

Konkordancja w Internecie

<http://www.webcorp.org.uk/>



The screenshot shows the WebCorp website interface. At the top, the logo "WebCorp" is displayed in a large, teal, 3D-style font. Below the logo is a navigation bar with several menu items: "WebCorp", "Advanced", "Wordlist Generator", "Guide", "Reviews", and "Feedback". A teal banner below the navigation bar contains a "NEWS" announcement: "Significant improvements in the speed of WebCorp and range of processing options available, as it becomes part of a fully-tailored linguistic search engine... [more]". The main search area is a light gray box containing a "Search term:" label, a text input field with the word "alquanto", and a prompt "Enter a word, phrase (no quotes necessary) or [pattern](#)". Below the search term field are three separate control boxes: "Search engine:" with a dropdown menu set to "Google", "Concordance Span:" with a dropdown menu set to "5" and the text "word(s) to left and right", and "Case options:" with a dropdown menu set to "Case Sensitive". At the bottom of the search area is a "Send Results by Email:" label. The entire interface is framed by a blue border with a vertical scrollbar on the right side.

WebCorp

WebCorp | **Advanced** | **Wordlist Generator** | **Guide** | **Reviews** | **Feedback**

NEWS: Significant improvements in the speed of WebCorp and range of processing options available, as it becomes part of a fully-tailored linguistic search engine... [more]

Search term:

alquanto

Enter a word, phrase (no quotes necessary) or [pattern](#)

Search engine: Google

Concordance Span: 5 word(s) to left and right

Case options: Case Sensitive

Send Results by Email:

Konkordancja w Internecie

KWiCFinder

<http://www.kwicfinder.com>

Aplikacja instalowana na komputerze,
przeszukiwanie Internetu.

Budowa własnego korpusu z Internetu

■ BootCat

- wyszukuje i zapisuje interesujące nas strony na podstawie podanych przez nas słów kluczy,
- dostępny na GNU General Public License.
 - Potrzebny nam dodatkowy program do obróbki tych tekstów.

■ Kolokacje A. Buczyński - program Kolokacje łączy w sobie robota internetowego i wyszukiwarkę kolokacji; GNU General Public License.

Sketch Engine

- WebBootCaT – tworzenie korpusu w oparciu o teksty internetowe i słowa kluczowe (*seed words*)
 - Konkordancja
 - Frekwencja
 - Kolokacje → *word sketch, common patterns*
-

CLAWS web-tagger

- Anotacja częściami mowy dla jęz. angielskiego
 - Użyty do anotacji BNC
 - Serwis internetowy
-

AntConc

- A freeware concordance program for Windows, Macintosh OS X, and Linux.
 - **WebParaNews** concordancer
 - A web-based interface to the Japanese-English News Articles database of the National Institute of Information and Communications Technology (NICT).
-