

Fictional Linguistics

[John Lawler](#)

Professor Emeritus of Linguistics
University of Michigan, Ann Arbor

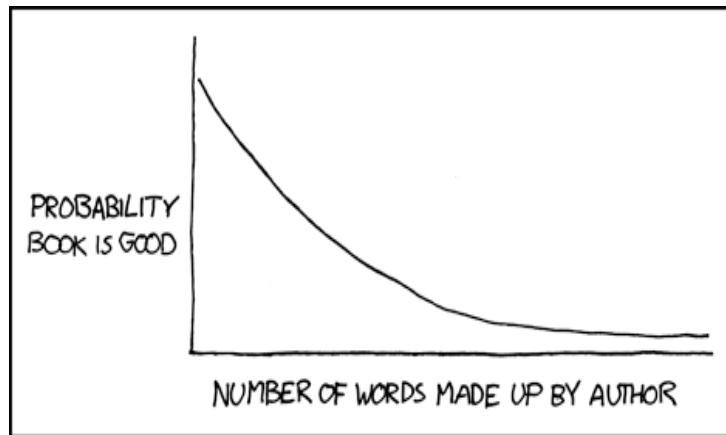
NorWesCon 34 2011

Fri 4/22 9pm Cascade 10

This handout available at <http://www.umich.edu/~jlawler/FictionalLinguistics.pdf>

Language in Science Fiction

<http://xkcd.com/483>



"THE ELDERS, OR FRAÁS, GUARDED THE FARMINGS (CHILDREN) WITH THEIR KRYTOSES, WHICH ARE LIKE SWORDS BUT AWESOMER..."

The Language Construction Kit

Online: <http://www.zompist.com/kit.html>

For a number of years, Mark Rosenfelder has maintained a lovely conlang web site with *The Language Construction Kit*, a small but crisp intro to linguistics for authors who want their linguistic science fiction to be scientific. I once said on the Web that it "contains all the information any SF writer would need to make up a **reasonable** alien language. And serves as quite a good introduction to linguistic concepts for beginners, to boot."

The Language Construction Kit is also a book by Mark Rosenfelder, [available from Amazon](#). It's much bigger and more inclusive than the online version. [Truth in Fictional Linguistics Dept: He asked me to blurb the book.] And this year he has a sequel, [*The Planet Construction Kit*](#), with geology, creatures, cultures, religions, technology, and war styles, plus how to make maps, illustrations and 3-D models.

A few links, to books, blogs, and people concerned with linguistic science (some in science fiction)

[The Cambridge Encyclopedia of Language](#), [The Cambridge Encyclopedia of the English Language](#)
(Both by [David Crystal](#), each 1-vol pb. Contain everything one needs to know about the subjects.)

Blogs: [Language Log](#), [Tenser, said the Tensor](#), [Language Hat](#), [Arnold Zwicky](#)

The best general online linguistic resources are the [LINGUIST List](#), the [Ethnologue](#), and the [Linguistic Term Glossary](#) (the last two are [SIL](#) productions). Most of the Wikipedia articles on linguistic terms and topics are reliable now (though stay away from peevings about minutiae of expression). [IPA](#) is useful.

[Suzette Haden Elgin](#): Linguist (PhD, UCSD) and SF writer (*Native Tongue*, Ozark trilogy, Coyote Jones series) publishes a newsletter about Linguistics in/and SF. Check out "[We Have Always Spoken Panglish](#)"

A list of Science Fiction Books featuring linguistics or linguists, [from Maggie Browning](#) at Princeton [How to Read Greek](#), or at least how to pronounce it and understand words (Ancient Greek, that is). [How to figure out a sentence](#), advice for readers (rather than writers, of course), and [an example](#).

The rest of the material in this brick-and-mortar handout represents a few of the many topics in linguistics that might pique the sense of wonder that any SF fan (and SF writer) has to have. There's lots more, and there's another talk Saturday at 3 here on historical linguistics (handout [here](#)). So please ask questions; only speak loudly, please, because I'm an old fart and wear a hearing aid.

Some phenomena of individual human languages

Lushootseed (**dxʷ-ləš-úcid** ‘our language’; Northern (Skagit) < **gʷídəq** ‘geoduck’, **scəqí?**, ‘sockeye’) is, like all languages of the Pacific Northwest, a **polysynthetic** language, i.e., most words are composed of a heavily inflected root, and most words are whole sentences. Lushootseed has seven different types of root reduplication, each with a different meaning. Three examples:

Lushootseed also has interrogative verbs, for instance:

gʷat '(to) be who?'	gʷat-ci?it 'Who is she?'
?əxíd '(to) be how?'	?əs-?əxíd-čəxʷ 'How are you?'
čad '(to) be where?'	čad-əxʷ-ti?it 'Where is he now?'
(?əs- 'temporary condition' - əxʷ 'now')	

Malay (Bahasa Indonesia, Bahasa Malaysia, Bahasa Melayu) has many interesting features:

Adjectives: **benar** ‘correct’ **kental** ‘thick (of liquid)’ **betul** ‘true’
 Derived nouns: **ke-benar-an** ‘correctness’ **ke-kental-an** ‘coagulation’ **ke-betul-an** ‘by coincidence’
 Liver, not ♥: **hati** ‘liver’ **ke-hati-hati-an** ‘care, caution’ **mem-per-hati-kan** ‘to pay attention to’
bilang ‘[1] to say; [2] every’ **Apa dia bilang?** ‘What did he say?’ **bilang hari** ‘every day’
jalan ‘[1] road, street; [2] to leave’ **Jalan Penang** ‘Penang Street’ **selamat jalan** ‘Goodbye’
matahari ‘(the) Sun’ [lit ‘eve of day’] **orang utan** [lit ‘jungle person’] **air batu** ‘ice’ [lit ‘rock water’]

Javanese, spoken in the most densely populated area in the world, has, not coincidentally, the world's most complex encoding of social status in its vocabulary. Almost every Javanese word comes in two or three versions, each faluting at a different social level. Thus, every Javanese sentence contains dozens of ways (many insulting or embarrassing) of commenting on the speaker's social relation with and opinion of the listener and everyone spoken of, simply by choosing different words. The only way to avoid making such comments is to avoid speaking, except in ritualistic formulas. An example, from p. 40 of *The Cambridge Encyclopedia of Language*:

Level	are	you	going	to eat	rice	and	cassava	now	Complete
krama inggil	menapa	pandjenengan	badé	dahar	kalijan	sekul	kaspé	samenika	Menapa pandjenengan badé dahar sekul kalijan kaspé samenika?
krama biasa									Menapa sampéjan badé neda sekul kalijan kaspé samenika?
madya	napa	sampéjan	adjeng	neda	sega	lan	saniki	Napa sampéjan adjeng neda sekul lan kaspé saniki?	Napa sampéjan adjeng neda sekul lan kaspé saniki?
ngoko madya									Apa sampéjan arep neda sega lan kaspé saiki?
ngoko biasa	apa	kowé	arep	mangan				saiki	Apa kowé arep mangan sega lan kaspé saiki?

Five status levels, in one Javanese dialect (after C. Geertz, 1968), using the sentence *Are you going to eat rice and cassava now?* The names *krama*, *madya*, and *ngoko* refer to 'high', 'middle', and 'low' respectively. In addition, the high and low levels each have two divisions, depending on whether honorific words are used, to produce *krama inggil* vs *krama biasa*, and *ngoko madya* vs *ngoko biasa*.

Some general phenomena of human language

Ergativity (found widely in Basque, Caucasian, Mayan, and Australian languages, among others)

In this system of grammatical relation marking, the “Object” of a **transitive** sentence (e.g., *They saw him*) and the “Subject” of an **intransitive** sentence (e.g., *He sat there*) are both marked the same way, in the **Absolutive** case, while the “Subject” of a **transitive** sentence gets marked differently, in the **Ergative** case; this means that in an Ergative system, there is no concept of Subject or Object. Most European languages use a different system (the **Accusative** system, named, like **Ergative**, for the special case that only occurs in transitive sentences) that marks Subject and Object, no matter what kind of verb they’re used with.

Most languages have some Ergative features; even English has a little of this in some compounds; everybody knows that *employer* and *employee* are complementary, but what about the *-ee* suffix in *transportee*, *parolee*, and *escapee*, for example? These are all Absolutive nouns, either subjects of intransitive verbs like *escape*, or objects of transitive verbs like *transport* or *parole* with indefinite subjects.

Noun Classification

There are several kinds of classifiers, usually displaying various parts of the same repertory of concepts and semantic categories in languages worldwide:

<u>Dimensionality</u>	<u>Animacy</u>	<u>Material State</u>
• 1-Dimensional	• Plant	• Solid / Fluid • Rigid / Flexible
• 2-Dimensional	• Animal	• Concave / Convex • Extended / Limited
• 3-Dimensional	• Human	• (Degree of) Angularity and Connectedness
• Paths and Combinations	• Body parts (usually human)	

Numeric classifiers are used with numbers – for instance, in Malay one cannot say **dua buku* (*dua* ‘two’ *buku* ‘book’) but instead must say *dua buah buku*. *Buah* ‘fruit’ is the classifier for non-animate nouns; *ikan* ‘tail’ is for non-human animals, and *orang* ‘person’ is for people. Malay is a very simple system, with only three classifiers. Japanese has more, including *hon* ‘1-Dimensional’, *mai* ‘2-D’, and *ko* ‘3-D’: *ni-hon no enpitsu* ‘two pencils’; *ni-mai no kami* ‘two sheets of paper’; *ni-ko no ringo* ‘two apples’; *ni-ko no nasu* ‘two [round] eggplants’; *ni-hon no nasu* ‘two [long thin] eggplants’. [Chinese](#) has a great many more (and Chinese speakers have a lot of fun fooling around with classifiers). [Burmese](#) has over 90, including specific classifiers for such things as betel quids, hands of bananas, lengths measured with a bamboo pole, and bundles of seedlings.

Classificatory affixes in an inflected language mark a transitive verb as having a particular sort of direct object. [Navajo](#), for example, has no single verb that means ‘to give’, but rather a verb root that must be inflected to mean giving a solid roundish object (ball, box, etc.), giving a load, pack, or burden (backpack, saddle, etc.), giving a slender flexible object (rope, socks, etc.), giving a flat flexible object (blanket, coat, etc.), and 8 more.

“Gender” is really a categorization system for nouns and concepts, allowing them to be tagged together (“gender agreement”), with sex usually irrelevant. English has no noun gender, Spanish and French have two genders, German and Latin three, Swedish a different three, and Polish four. Bantu languages, however, typically have ten to twenty noun classes, sorted not by sex but by meaning category. Examples from [Swahili](#):

Trees: [sg] *m-limau*, [pl] *mi-limau* ‘lemon tree(s)’; *m-ti*, *mi-ti* ‘tree’; *m-witu*, *mi-itu* ‘forest’

Fruit: *limau*, *ma-limau* ‘lemon’; *pera*, *ma-pera* ‘guava’ People: *m-tu*, *wa-tu* ‘person’;

m-zee, *wa-zee* ‘old person’; *m-toto*, *wa-toto* ‘child’; *m-wana*, *wa-ana* ‘son/daughter’

“Things”: *ki-ti*, *vi-ti* ‘branch’; *ki-tabu*, *vi-tabu* ‘book’; *ki-toto*, *vi-toto* ‘infant’; *ki-su*, *vi-su* ‘knife’; *ki-kapu*, *vi-kapu* ‘basket’; Abstractions: *ukubwa* ‘size’, *umoja* ‘unity’, *uzee* ‘old age’

Na’vi Phonetics and Phonology

Na’vi has 20 consonants, 7 vowels, 4 diphthongs, and 2 syllabic “pseudovowels,” rr and //. Na’vi has a 7-vowel system:

i , ì	u
e	o
ä	a

Diphthongs: aw [aw], ew [ɛw], ay [aj], ey [ɛj] (Notice: no oy, ow, äy, äw])

i	[i]
ì	[I]
e	[ɛ] Note: always lax
ä	[æ]
u	[u] or [U]
o	[o]
a	[a]

Ejectives:

px [p’] **t**x [t’]

kx [k’]

Voiceless Stops:

p

k

[?]

Affricate:

ts

Voiceless fricatives:

f

s

h

Voiced fricatives:

v

n

ng [ŋ]

Nasals:

m

r, l

Liquids:

w

y [j]

Glides:

And in the news...

From: <http://languagelog.ldc.upenn.edu/nll/?p=1977>, by Paul Frommer

Ranks in Michigan’s ‘Hutaree Militia’ (from [Language Log](#))

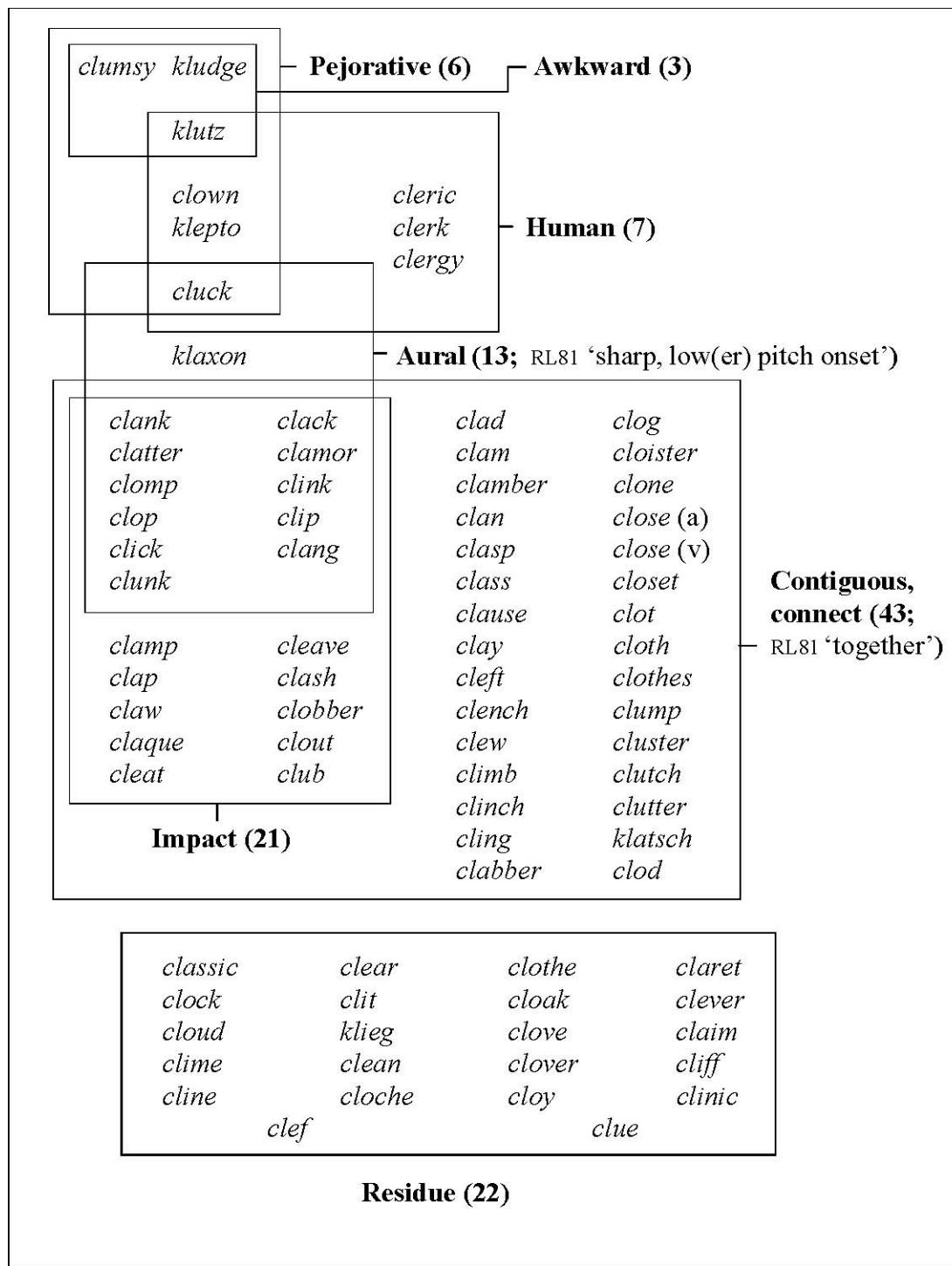
RADOK [RD]
 BORAMANDER [BM]
 ZULIF [ZL]
 ARKON [AK]
 GOLD RIFLEMAN [GR]
 SILVER RIFLEMAN [SR]
 BRONZE RIFLEMAN [BR]
 LUKORE [LK]
 MASTER GUNNER [MG]
 SENIOR GUNNER [SG]
 GUNNER [GN]

(below is a comment by Nick Lamb from the LL post above)

This chart has 12 ranks, and I doubt they have more than a few hundred members. To me that says it's functioning more like the ranks in a treehouse club. The kid with the treehouse gets to be "Radok" and his best friend is a "Boramander". The kid they don't like much is a "Lukore", except in the summer when they use his parents' pool, then he's "Gold Rifleman" and nobody is ever a "Silver Rifleman" because the Boromander's sister took the silver pen.

We should be able to do better than this.

Sound symbolism – words whose meaning is related to their sound – is alive and well in English, as well as other languages. For instance ...



84 Simplex Words with Assonance KL-

Coherence 62/84 = 74%

From J. Lawler 2009. “[The Data Fetishist’s Guide to Assonance Coherence](#)”
(RL81 = Rhodes & Lawler 1981. “[Athematic Metaphors](#)”, CLS 17)

A bit of syntax. The details of English Negative Polarity phenomena are a good example of how ridiculously complex any real language can get. It’s all a matter of combinations, like [this problem](#).

RC HUMS/LING 393

English Grammar and Writing

Negation & NPIs

Negative Polarity Items

(illustrated with *not*)I *(don’t) have **any**.[but **NOT** *any* as Subject, e.g., *Anyone can do that*, which is Modal Polarity.]I *(don’t) **ever** see him. [*ever* = **anywhen*]He would*(n’t) **budge**.He did*(n’t) have a **red cent**.

He has*(n’t) been here

in weeks/ages/the longest time.He may *(not) **arrive until** noon.[*until* with punctual predicates]I *(don’t) have **much** time left.I *(don’t) have **many** days left.He has*(n’t) arrived **yet**.You **need***(n’t) stare at him.We **dare** *(not) disturb him.[*need & dare* as modals]I *(don’t) like that **at all**.It’s *(not) that easy **any more**.He’s *(not) **all that smart**.[*be all that + adjective/adverb*]He’s *(not) **too** bright. [*too* = *very*]*(Don’t) **bother to** close/closing the door.That should*(n’t) **last long, be long,****take long.** [but **NOT** *a long time*]I **can***('t) **seem to** understand him.I **would***(n’t) **care to** fight with her.I **would***(n’t) **mind** fighting with her.I *(don’t) know **but that/what** he’s right.They *(don’t) **drink a drop, do a thing,****give a damn/shit, lift a finger,****bat an eye, eat a bite**, ... [V+minimal D.O.]They **can***('t) **help** themselves.They **can***('t) **help** thinking of that.I saw *(no) people there **to speak of**.

* Asterisk before a sentence indicates ungrammaticality.
 Parentheses indicate optional material that can be omitted.
 Asterisk **before** parenthesis indicates that the elements inside the parentheses are obligatory and **cannot** be omitted.
 Asterisk **after** parenthesis indicates that the elements inside the parentheses are ungrammatical and **must** be omitted.

Negative ‘Triggers’

(illustrated with *any*)

I. Overt negatives:

A. **not** [w/ NPI clausemate or complement]He did*(n’t) know **anybody**.He did*(n’t) claim (that) he knew **anybody**.He did*(n’t) think (that) he knew **anybody**.

B. Incorporated negatives

I **doubt** (that) he knows **anybody**.It’s **unlikely** (that) he knows **anybody**.It’s **improbable** (that) he knows **anybody**.He **dislikes** **anybody**(’s) reminding him.He **prevented** her (from) seeing **anyone**.I **kept** her **from** telling **anybody**.

C. Negative frequency adverbs

I **seldom/rarely** see **any** of them.He **hardly/scarcely** knows **anyone**.

D. Quantifiers & quantified adverbs

Only Bill did **any** of the homework.(*A) **Few** people see **any** use for it.

II. Questions (overt and embedded)

Did you see **anybody**?How does **anybody** stand this?I wonder how **anybody** stands this.

III. Hypothetical clauses

Tell me **if** you see **anyone**.He asked **whether** I saw **anyone**.Tell me **whatever** **anybody** says.Check the list again, **lest** we forget **anyone**.I’ll read it aloud, **unless** **anyone** objects.

IV. Comparatives, superlatives, etc.

There’s **more/less** here than **anybody** knew.He **prefers** beer to **any** other drink.He’d **rather** die than hurt **anyone**.He’s **as** good **as** **anyone** expected.He’s the **fastest** one (that) **anybody** knows of.I saw him **before** **anybody** (did).I’m **surprised** (that) he knows **anybody**.It’s **too** dark to recognize **anything**.He left **without** **anybody** noticing (it).That’s **hard** for **anybody** to do.