

# ENGLISH AND EASTERN PENAN: A LEXICAL COMPARISON

Ian Mackenzie

*last amended 12 May 2018*

(Parts 1, 2 and 3 were originally posted in 2012 as an online appendix to

Mackenzie, Ian (2012). *Lexicography among hunter gatherers: the ECD as an ethnographic tool*. In: *Smysly, teksty i drugie zahvatyvayuschie syuzhety / Meanings, texts, and other exciting things: a Festschrift to Commemorate the 80th Anniversary of Professor Igor Alexandrovič Mel'čuk*

Note that the experiment was refined and amended in 2017, yielding the results shown below; these are slightly different from the results reported in the above-cited article.)

## 1. Introduction

The author has been conducting linguistic and ethnographic research among the Eastern Penan people of Sarawak, Malaysia, since the mid-1990s. Part of his work has been published in the form of a provisional dictionary of their language; the 2006 version of the dictionary can be downloaded gratis at

<http://www.rimba.com/pdindexf/pdindex.html>

The following presents an experiment comparing the lexicons of English and Eastern Penan. The object is to determine to what degree English lexemes and phrasemes can be readily translated into Penan. A random sample is taken from an English dictionary, and an attempt made to write definitions in Penan for each entry. As a control, the experiment is repeated, this time with French as the target language. The results are presented below, and the relevant conclusions drawn. Also discussed is the degree to which the Penan lexicon can be readily translated into English.

## 2. English to Penan Lexical Translatability

### 2.1 The Sample

The following English lexemes and phrasemes are all listed on p 340 of the Longman Active Study Dictionary of English, twentieth impression, 1991. The page was chosen at random. Only minor changes to the list were made, e.g. eliminating lexemes the author doesn't personally use, and removing redundancies arising from the existence of

different parts of speech denoting one and the same semantic unit. (Thus, while retaining *lanky* as the principal form, the derivative forms *lankily* and *lankiness* were discarded.)

I have refrained from adding lexemes, even though experience elsewhere tells me that by searching for word senses overlooked by the lexicographers, I could easily double the number of entries. However, most such new lexemes would be very close semantically to those already listed. To save time I have refrained from carrying out such an expansion, but I fully recognize that this failing makes it *easier* to find Penan equivalents for English words. That's because as things stand, I only have to find a single Penan equivalent for any English vocable which, although listed as a single lexeme, in reality has two or more senses. If Penan has a means of easily expressing one of those senses but not the other(s), then the language is semantically deficient in relation to English; but this deficiency will not register in the data. Thus, the general lexical mismatch between English and Penan that our little experiment reveals is likely even greater than it seems.

I often replace or improve the original English definition given in the dictionary; my goal is not to write an ECD<sup>1</sup> definition, or even a reasonably good informal one, but merely to make clear to a native English speaker exactly what sense is intended.

In some cases I have eliminated the examples given by Longman, and replaced them by others which I find more suitable. Like the examples in the original, these are italicized.

When a given vocable represents more than one lexeme, each of these is assigned a distinct number, e.g. landslide 1, landslide 2.

My definitions or explanations in the Penan language resemble the explanations I would use were I with real Penans and seeking to communicate the meanings to them as efficiently as possible. They are not always rigorous definitions in the ECD sense. I have retranslated them back into English as literally as possible.

### 2.11 Symbols used

A lexeme that is translatable by a Penan lexical term that is synonymous or nearly synonymous is unmarked, but the Penan equivalent is capitalized -- see the first example. Otherwise, the lexeme is marked with one of the three following symbols:

- √ easily translatable by means of a synonymous or nearly synonymous non-lexicalized multiword expression, or by means of listing two or more Penan lexemes that are narrower synonyms of the English
- ✂ requiring an expression that exceeds eighteen words or so which, despite its length, is an accurate translation
- \* no practical or possible means of translation; in other words, this category covers two kinds of case: (1) where a paraphrase might need to be paragraphs or pages in length, and (2) cases of *absolute*

---

<sup>1</sup> Explanatory-Combinatorial Dictionary; for an explanation, see the Introduction of the Eastern Penan Dictionary at <http://www.rimba.com/pdindexf/pdindex.html>

*untranslatability*, i.e. meanings that cannot be rendered into Penan no matter how long the proposed paraphrase. I do not distinguish the two cases here, because to do so accurately in any given case would require a level of semantic analysis that would amount to a research project in its own right. Nonetheless, a probable example of (1) is **21**; and a probable example of (2) is **23**. To define the latter in Penan almost certainly requires the use of semantic primes which cannot be expressed in Penan.

## 2.12 The Corpus

- 1 landslide 1, n. -- earth movement || **BESALÉ**
- 2 \* landslide 2, n. -- overwhelming victory in an election || (no word for "election" in traditional Penan; no word for "government" or "parliament either". Could likely explain this with a long paragraph.)
- 3 √ lane 1, n. -- small road || *jalan kerita éh si'ik* 'a narrow motor vehicle road'
- 4 ☒ lane 2, n. -- alley (small road between two main streets in a city) || *jalan kerita éh si'ik belah duah jalan kerita éh ja'au; jalan kerita éh si'ik nah ngepata duah jalan ja'au nah; boh pu'un pina lamin belah jalan ja'au nah ngan jalan si'ik nah* 'a narrow motor vehicle road between two wide motor vehicle roads; and the narrow motor vehicle road runs parallel to the two big roads; and there are many houses between the big roads and the little road' (in practice, one would just end up drawing a sketch)
- 5 \* lane 3, n. --division of a wide road || Could conceivably use *jalan* which is a general word for 'place where something is, or along which something moves', but only in strict context, because *jalan* also denotes the whole road! In any case, Penans only see dirt logging roads. But a long explanation could communicate this concept.
- 6 \* lane 4, n. -- running track in sporting event || *jalan* (preceding remark partly applies to this case. The whole track would also be a *jalan*.)
- 7 \* lane 5, n. -- fixed path for ships || In practice impossible to explain to traditional Penans, who think the ocean is a big river; in principle could explain with a long paraphrase.
- 8 language 1, n. -- human speech || **HA'**
- 9 language 2, n. -- subset of language 1, i.e. language used by specialists: e.g. language of science <auto mechanics> || **HA'** (But note that you couldn't translate the first example, because the concept "science" probably cannot be explained in Penan due to the absence in that language of certain semantic primes. But you could translate the second as: *ha' éh dilem éh penakai irah tukeng petenep kerita* 'the deep language used by the experts who fix cars')
- 10 \* language 3, n. -- a figurative sense of language 1, roughly speaking, semiotics; e.g. language of music <sculpture> || Impossible: the primary sense of *ha'* is 'sound', so this word can only denote audible phenomena. In regard to the first example, there is no general word for "music" -- traditionally Penans did not sing, and since they have no general word for "instrument", you have to be specific and say e.g. *ha' keringot* 'sound [of] noseflute'. Thus, even if you could get them to understand the meaning of the English word *music*, and then got them to borrow it into their language (both are possible), *ha' music* would just mean 'the sound of music'. As for the second example, translation is doubly difficult: sculpture is a branch of art, and there is no word or concept of "art". Note that taking a different tack and trying to define *semiotics* would very likely be impossible in its own right.

- 11 ☒ language laboratory, n. || *jah luvang lem sekolah ja'au (tong lebo kapan) jalan anak omok menéng avé pekalai ha' lia éh nala guru' réh lem ijin rikot* 'a room in a large school (in a town) where children listen to and practice foreign languages that have been placed in a "record" engine by their teachers'
- 12 √ languish, v. -- X stays for too long in a place where X should not be and does not deserve to be *He languished in that foreign land for ten years.* || *X moko peseng lebé lem jah retek éh bé' payo atau bé' kenelo X* 'X is stuck for a long time in a place that is not suitable for X or that X does not like'
- 13 √ languor, n. -- person's lack of vigour or vitality || *X bé' tigéh atau gahang kenin juk maneu ineu ineu* 'X is not strong/energetic, or has little will for doing anything at all'
- 14 √ lanky, adj. || *X éh kebit ngan mago* 'X who is tall and thin'
- 15 √ lantern, n. || *lapung éh ngelayau pu'un rang éh omok nihin ngan usah* 'a lamp that normally has a hoop handle and can be carried with one'
- 16 lap 1, n. -- place on thighs of seated person || **KONYONG**
- 17 √ lap 2, n. -- single journey round a track || *jah liwet sio réh petaké ngeliwet padang* 'one round when people are racing round a sport field'
- 18 ☒ lap 3, v. -- X laps Y = X passes a competing racer Y and moves one lap 2 ahead of Y || *X ngelépéng Y lem petaké, boh X lakau rigah kepéh ngeliwet padang kura kura kolé kepéh boh tong ga' néh X ngelépéng Y keruah kolé.* 'X passes Y in a race, and then X keeps on moving fast in the circuit round the field, any number of times, until X passes Y a second time'
- 19 √ lap 4, v. -- drink with quick movements of tongue || *mesep pakai jela' nyavu gaya' aseu atau méu* 'drink with the tongue as if imitating a dog or cat'
- 20 ☒ lap 5, v -- lap against -- water repeatedly gently impacts something solid || *lipak si'ik si'ik teneng tong dirin ba atau tong barei alut bé' ma'o ma'o maneu ha' si'ik si'ik.* 'very small waves strike against the shore or against things like boats again and again making small noises'
- 21 \* lapel, n. || In principle this could be paraphrased in a long paragraph, but since one would have to begin by describing what a jacket is, in practice one would just draw a picture.
- 22 ☒ 1. lapse, n. -- X has a small failure due to momentary lack of attention that results in faulty speech or action *Don't blame him for his lapses; he's old.* || *Uban X telalau metok, X salep atau sala' lem penganeu néh atau lem ha' néh éh nané néh.* 'Because X becomes distracted for a moment, X does or says something incorrectly'
- 23 \* lapse 2, v -- change to a state of inaction or non-existence due to passage of time. *That provision of the law will lapse in 2015* || There is an easy way of saying 'over' or 'terminated', but this English lexeme is more complex than that, and also seems to apply only to phenomena that Penans do not understand or have words for -- like contracts and laws. To define these, as well as the rough definition just offered, is likely impossible due to the lack of certain primes.
- 24 \* lapse 3, v. -- cease to exist due to failure to exercise timely action *If you do nothing, your rights will lapse* || see comment immediately above
- 25 ☒ lapsed, adj. -- X is a lapsed Y = X has permanently stopped practicing religion Y *She is a lapsed Catholic* || *X lebé sebayang kivu gaya' sebayang Y, boh tong ga' néh X ma'o sebayang Y, boh X bé' jam sebayang Y kepéh* 'for a long while X prayed in the manner of Y prayer, but in the end X stopped performing Y prayer, and X is unlikely to perform Y prayer again'

- 26 larceny, n. || **NEKAU** (this is the verb meaning 'steal')
- 27 √larch, n. -- kind of tree || **jah bengesa' kayeu** 'a kind of tree'
- 28 √lard, n. || **nyak babui atau idok éh pepunyai reh boh meket kepéh** 'wild-pig or domestic-pig fat that someone has melted and that has then congealed'
- 29 √larder, n. || **luvang si'ik mu'un éh pu'un kep éh jalan réh modo penguman éh juk kinan réh da'ap** 'a small enclosed space with a lid/door that people use to store food that they are going to eat soon'
- 30 1. large, adj. || **JA'AU**
- 31 √large: X is at large. -- X is free or has escaped, and poses a danger to someone || **X pelepu atau lakau sé jam semah, boh X omok maneu pengesa'at.** 'X has escaped from confinement or is moving around who knows where, and X can do harm'
- 32 \* largely, adv. *The island is largely undeveloped; This claim is largely accepted.* || Impossible to paraphrase elegantly, for there is no word meaning anything like 'largely,' 'most', 'mostly'. However, although the meaning cannot be generally defined, any given example in which it occurs can be communicated by a particular kind of circumlocution. For example, one would translate the first example as *Pu'un pulau tong ba banget. Si'ik awah tana' tong pulau inah éh pu'un lamin tong néh.* 'There is an island in the River Ocean. (there are no lakes in Penan territory, and the word for "island" -- borrowed from Malay -- normally covers an unlogged patch surrounded by destroyed forest.) Only a small amount of the land on that island has houses on it.' (there is no word for 'developed'). The second example would have to be rendered, *Pina kelunan ngelan tong ha' néh, bé' pina kelunan éh bé' ngelan tong ha' néh.* 'Many people rely on what he/she says, not many people do not rely on what he/she says'.
- 33 √largesse, n. || **Ineu ineu éh nena' irah éh kaya' asan néh jin kenin réh awah.** 'Anything given out by people who have things in abundance, as long as they give it from their hearts and not because they have to'
- 34 ☞ lark: for a lark -- have a bit of fun that isn't quite the normal thing you do for fun, and which does no real good but causes no real harm either || **seminga' metok, inga' ko' tengé jin inga' ko' éh ngalayau, sio ko' seminga' nah bé' pu'un ineu ineu éh gunah naneu ko', tapi' bé' pu'un pengesa'at naneu ko'** 'play for a short while, your playing not being usual for you, without doing anything useful, but without doing anything bad'
- 35 √lark, n. -- kind of bird || **jah bengesa' juhit** 'a kind of bird'
- 36 ☞ larva, n. || **ulet éh jam malui ka'an éh ja'au jin néh, layan néh tengé mu'un jin néh, barei langau, besungah, legang, layuk, avé éh jah jah péh** 'grub/caterpillar that can transform into a creature larger than itself, that looks very different from itself, such as a fly, a *Rhynchophorus ferrugineus* beetle, a butterfly/moth, a honey bee, or many other things' (note that this definition might cause surprise, because the Penans are ignorant of the life cycles of certain insects, e.g. butterflies, and in fact believe that hairy caterpillars can turn into squirrels)
- 37 √laryngitis, n. || **sakit lem tegoro' maneu ka'au tusah atau sakit hun ko' juk pané** 'a sore throat that makes it hard or painful when you want to speak'
- 38 √larynx, n. || **tegoro'** (this Penan lexeme actually denotes the adam's apple, or, in the case of women and children, the place where the adam's apple would be)

- 39 lascivious, adj. || **LO'O**
- 40 \* laser, n. -- a definition might read "device that emits light through a process of optical amplification based on the stimulated emission of electromagnetic radiation, such that the photons so emitted are coherent (= the crests and troughs of the light waves are in lockstep, in contrast to ordinary light); and such that they are all of one wavelength (one specific colour); and such that they are in one tight beam that stays focused for vast distances". || A definition along these lines would require terms that do not exist and cannot exist in Penan due to the absence of certain semantic primes (preventing the translation of e.g. *radiation*, *wavelength*, and *colour*.) However, this is a technical definition, and is not the sense present in the minds of the average English speaker, which is more like 'a special device that emits a certain kind of light, consisting of a beam that does not grow wider no matter the distance from the device, and can be made so intense it can blind or burn'. This latter definition could be at least roughly translated; perhaps the biggest problem is the lack of a word for 'device'. Instead, *livah* could be used -- this means 'cloth' but can have the extended meaning of a thing made or owned for a purpose by a human being. There is no Penan word for 'beam', but this could be explained with a simile, e.g. like the shape of a blowdart shaft.
- 41 lash 1, v. -- whip || **MEMERAH**
- 42 √ lash 2, v. -- tie together sturdily with cord || *ngejeret*, *negaleu*, *milit* (narrower synonyms -- each of these refers to different kinds of tying or lashing)
- 43 √ lash out 1, v. -- attack violently and rapidly || *beleka' nyurung* 'suddenly attack'
- 44 √ lash out 2, v. -- lash out 1 figuratively, with words, due to sudden anger || *pakai ha' mahéng uban néh beleka' merek* 'use harsh words because he/she suddenly angry' (note that such behaviour violates a strong prohibition in Penan culture)
- 45 ✂ lasso || *labak éh penakai kelunan éh juk molong kuda avé sapi' éh penakai réh hun réh nekelak ka'an inah boh réh ngaléng labak nah kenéh teneng tong batok ka'an nah jalan néh jadi telun néh jalan réh omok ngamit éh* 'snare noose used by people who want to keep horses and cows as pets and used by them when they chase after the animals and they throw the noose so that it catches on the neck of the animal and becomes the animals' leash, letting them lay their hands on it'
- 46 last, n, pron. -- ultimate element in a sequence || **GA'**

## 2.2 Summary

All English entries, classified according to ease of translation into Penan:

Translatable by a Penan lexical item that is synonymous or nearly synonymous		9
Translatable by a synonymous or nearly synonymous non-lexicalized multiword expression, or by means of listing two or more narrower synonyms	√	18
Translatable by a synonymous or nearly synonymous multiword expression, but exceeding 18 words	✂	9
No practical or possible means of translation	*	10

### 2.3 Conclusion

There are 46 English lexemes and phrasemes listed here. As (2.2) shows, in only 9 cases do Penan lexemes exist with meanings sufficiently close to the English ones to be given as simple equivalents. In 18 other cases, the Penan language has the ability to construct satisfactory paraphrases of the English lexical items. (Note that two of these -- (27) and (35) -- name species that do not exist in Borneo. Their meanings in Penan are 'a kind of tree' and 'a kind of bird', respectively. From a purely lexicological point of view, these definitions are adequate. Descriptions of these species, which may include pictures, belong to the encyclopaedic zone of the dictionary. This latter information is non-linguistic in nature.)

However, 19 of the English lexical items, that is more than 40% of them, defy translation into Penan to one degree or another. They require translations that are excessively long, imprecise, or both. Ten of the English lexemes are in practice untranslatable; some of these are doubtless untranslatable in theory, insofar as they would require definitions that make use of semantic primitives that are not expressed in the Penan lexicon.

## 3. English to French Lexical Translatability

### 3.1 A Control Experiment

It is instructive to compare this set of (attempted) translations from English to Penan with one from English to another language. For this comparison I have chosen French, largely for convenience; my competence in the latter language is close to that of a native speaker. French also has the advantage of sharing much of the history and culture of English, and thus provides a good contrast to a language spoken by isolated hunter gatherers.

In what follows I take the English lexemes and phrasemes that appear in (2.12) and attempt to translate them into French.

Where an existing French lexeme or phraseme can be used directly as a gloss, it is given in non-italic type. Where the French language has no perfect lexical equivalent for the English item, and recourse must be made to a definition or a list of partial synonyms, the expression(s) used for translation is (are) given in italics.

### 3.2 The Corpus

landslide 1, n. || **glissement de terrain**

landslide 2, n. -- overwhelming victory in an election || *victoire écrasante lors d'une élection*

lane 1, n. -- small road || **chemin, petite route**

lane 2, n. -- alley (small road between two main streets in a city) || **ruelle**

lane 3, n. -- division of a wide road || **voie**

lane 4, n. -- running track || **couloir**

lane 5, n. -- fixed path for ships || **couloir de navigation**

language 1, n. -- human speech || **langue**

language 2, n. -- subset of language1, e.g. language of science <auto mechanics> || **langage**

language 3, n. -- figurative sense of language1, e.g. language of music, mathematics || **langage**

language laboratory, n. || **laboratoire de langues**

languish, v. -- X stays for too long in a place where X should not be and does not deserve to be *He languished in that foreign land for ten years.* || **languir**

languor, n. -- person's lack of vigour or vitality || **languueur**

lanky, adj. || **longiligne**

lantern, n. || **lanterne**

lap 1, n. -- place on thighs of seated person || **genoux**

lap 2, n. -- single journey around a track || **tour de piste, tour de circuit**

lap 3, v. -- pass a competing racer and move one lap2 ahead || **avoir un tour d'avance**

lap 4, v. -- drink with quick movements of tongue || **laper**

lap 5, v -- lap against -- water repeatedly gently impacts something solid || **clapoter**

lapel, n. || **revers**

lapse 1, n. -- small failure due to momentary lack of attention *Don't blame him for his lapses; he's old.* || **défaillance, trou de mémoire, relâchement de l'attention**

lapse 2, v -- change to a state of inaction or non-existence due to passage of time. *That provision of the law will lapse in 2015* || **être caduc ou périmé**

lapse 3, v. -- cease to exist due to failure to exercise timely action *If you do nothing, your rights will lapse* || **tomber en désuétude**



lapsed, adj. -- X is a lapsed Y = X has permanently stopped practicing the Y religion *She is a lapsed Catholic* || **pas pratiquant** (*i.e. X est un Y qui n'est plus pratiquant*)

larceny, n. || **vol**

larch, n. -- kind of tree || **mélèze**

lard, n. || **saindoux**

larder, n. || **garde-manger**

large, adj. || **grand**

large -- X is at large = X is free or has escaped, and poses a danger to someone || **en fuite, en liberté**

largely, adv. *The island is largely undeveloped; This claim is largely accepted.* || **en grande partie**

largesse, n. || **largesse**

lark: for a lark -- have a bit of fun || **pour rigoler**

lark, n. -- kind of bird || **alouette**

larva, n. || **larve**

laryngitis, n. || **laryngite**

larynx, n. || **larynx**

lascivious, adj. || **lubrique**

laser, n. || **laser**

lash 1, v. -- whip || **fouetter**

lash 2, v. -- tie together || **attacher ou fixer fermement à l'aide d'une corde**

lash out 1, v. -- attack suddenly and violently || **devenir violent, se démener**

lash out 2, v. -- lash out1 figuratively, with words || **invectiver**

lasso || **lasso**

last, n, pron. -- ultimate element in a sequence || **dernier**

### 3.3 Summary and Conclusion

There are 46 English lexemes and phrasemes. For 41 of these, there exist French lemmemes or phrasemes sufficiently close to the English items to be given as simple equivalents. In only five cases am I forced to use non-lexicalized French expressions, or a list of partial synonyms, to describe the meanings of these English lexical items, and in each case these paraphrases are short and sufficiently accurate.

The result of the lexical comparison in (3.2) is strikingly different from that revealed in (2.12). The semantic differences between English and French are almost trivial when compared with the profound differences that divide English and Penan. It would be instructive to carry out many other surveys of this nature, comparing the lexicons of a wide selection of the world's languages.

## 4. Penan to English Lexical Translatability

Of course, a third experiment is called for: a random sampling of Penan lexemes, followed by the attempt to find English equivalents. Unfortunately, this cannot be performed with the same reliability as the above two experiments, for there does not yet exist a Penan dictionary of the requisite length. A decent English dictionary contains many tens of thousands of word meanings and idioms, but even the latest draft (as yet unpublished) of the Penan dictionary has only thirteen thousand entries. This is by no means the extent of Penan vocabulary; rather, evidence, if any is needed, of the enormous challenge of lexicography. There is no reason to believe that a Penan's vocabulary is much smaller than that of an English speaker; we are talking of a lexicon that may exceed a hundred thousand items.

Still, despite the limitations of the current Penan dictionary, it remains the best source available. As it turns out, only about a quarter of the lexemes listed in it have a reasonably close single-word English equivalent. Three quarters of the entries require a definition containing a number of words. And in about a third of these -- a quarter of the total -- the definitions can be long, and often complex and subtle. (For example, see the semantic analysis of *pelin* at <http://www.rimba.com/pdlextestf/pdlextest.html>.)