

Remarks on John Haiman, 1999.  
'Auxiliation in Khmer: the case of *baan*'  
*Studies in Language* 23:1. 149–172

N.J. Enfield

Max Planck Institute for Psycholinguistics, Nijmegen

Haiman's article on the polyfunctional word *baan* in Khmer touches on a fascinating area in Southeast Asian linguistics, with conclusions of theoretical importance to the study of grammaticalisation and of universals in linguistic typology and linguistic change. However, there are serious problems with Haiman's presentation. I would like to point out why some assumptions Haiman makes are problematic, and why the merits of his concluding speculations cannot yet be assessed. (For a detailed treatment of Khmer *baan* and its analogue in neighbouring languages, see Enfield 2000, in press.)

Haiman begins by stating that

the most striking — and perhaps the only noteworthy — feature of the auxiliation of the main verb *baan* 'get' in Khmer is that it migrates from V2 to V1 position, contravening the general tendency for grammatical morphemes to remain frozen in the same position where the words from which they originate are found.

Apart from the fact that there are many noteworthy features of the polyfunctionality of *baan*, there remain two highly problematic ideas in Haiman's opening remark. The first is the unquestioned assumption that there has been such a 'migration' at all, despite possible alternative explanations (e.g. that both pre- and post-verbal *baan* derived separately and directly from a main verb related to present-day *baan* 'get'). Nowhere does Haiman consider alternative accounts for the emergence of modern 'auxiliary' functions (both pre- and post-verbal) of *baan*. Second, there is an error of temporal logic embodied in the statement that morphemes may or may not 'remain frozen in the same position where the words from which they originate *are* found' (my emphasis). The words from

which these morphemes originate are *not* ‘found’ in the same data set as the said morphemes. The source morphemes are of a former era, not of the present. And nowhere does Haiman discuss historical data. The relatedness Haiman seeks to illustrate is restricted entirely to analysis based on contemporaneous forms. While synchronic relatedness between various meanings of polysemous contemporary forms is relevant to hypotheses about historical change (especially with respect to pragmatic processes of consistent and repetitive inference leading to emergence of new meanings), we are not entitled to automatically regard synchronic relatedness (however that is established) as evidence of the presence or nature of diachronic relatedness. (It should be added that Haiman’s contribution to the study of grammaticalisation is not unusual in making the unwarranted leap from synchronic to diachronic relatedness.)

Haiman supplies contemporary examples of *baan* in various functions, and states that there are three morphemes, one a main verb, one a post-verb (or V2), one a pre-verb (or V1). (See references in Haiman’s paper for details on the functions of *baan*.) Here, for purposes of discussion, I will refer to the basic three patterns that Haiman identifies, as follows:

1. **Main-V *baan***: NP1 *baan* NP2 ‘NP1 “got” NP2’
2. **Post-V *baan***: NP1 VP *baan* ‘NP1 can/was able to VP’; ‘NP1 VP-ed “with success”’
3. **Pre-V *baan***: NP1 *baan* VP ‘NP “got/managed/happened” to VP’; ‘NP1 “did” VP’

Figure 1. Three functions of *baan* in contemporary Khmer

(Note that there are other, important, and grammatically distinct functions of *baan* — for example in heading adverbial adjuncts of extent and manner — which Haiman does not mention.) To make more explicit a logical problem in Haiman’s argument, it would be accurate to explicitly label the three *baan*’s in Figure 1 as ‘Today’s main-V *baan*’, ‘Today’s post-V *baan*’ and ‘Today’s pre-V *baan*’, since Haiman’s description is based on contemporary data. Haiman posits historical development from one *baan* to the next, assuming an *in-series* path of grammaticalisation as follows:



Figure 2. Haiman’s hypothesised linear grammaticalisation path

Haiman then seeks to account for a putative leftward ‘migration’ of *baan* (i.e. from post-verbal ‘auxiliary’ to pre-verbal ‘auxiliary’), in step 2 > 3 of Figure 2. Haiman cannot mean that today’s pre-V *baan* developed from *today’s* post-V *baan*,

which before that had developed out of *today's* main-V *baan* (although that is how he presents it). Making the historical process explicit (i.e. by explicitly recognising that the sources of today's forms are the forms of a previous age), let's assume Haiman means that today's pre-V *baan* developed (via his 'migration') from a predecessor of today's post-V *baan*, which, before that, had developed from a predecessor of today's main-V *baan*, as follows:

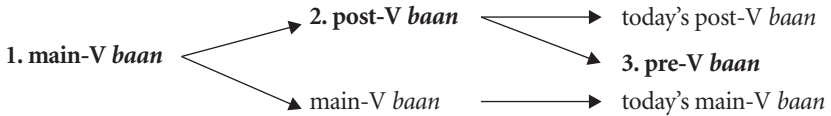


Figure 3. A more explicit representation of Haiman's hypothesis

An initial problem here is the assumption that the post-V *baan* of stage 2, and the main-V *baan* of stage 1 have the same semantic/grammatical properties as today's main- and post-V *baan*, since the information we are given on these putative historical source morphemes is drawn only from contemporary data. Thus, a serious gap in the argument appears, but empirical evidence from past stages of the language would verify or falsify the claim. For the sake of argument, let's adopt Haiman's assumption that today's post- and main-V *baan* are identical in the relevant respects to the post- and main-V *baan* of former stages. This would validate Haiman's claim of historical relatedness between contemporaneous data. But now a new problem emerges. Why should it be assumed that today's pre-V *baan* emerged from a post-V *baan*, and not from a main-V *baan*? Haiman does not mention the following alternative:

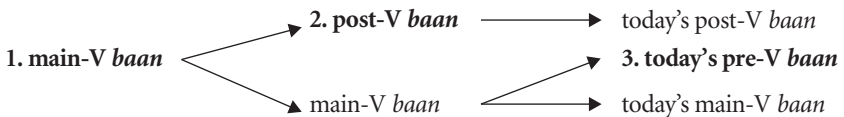


Figure 4. An alternative to Haiman's 'migration' hypothesis

or, more simply:

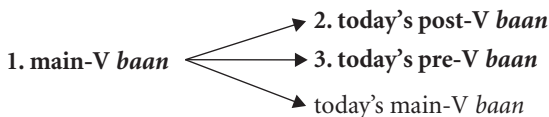


Figure 5. Another alternative to Haiman's 'migration' hypothesis

Indeed, it may also have been that today's pre-V *baan* appeared prior to post-V *baan*, which hypothesis, again, could be verified or falsified by empirical evidence.

(It is notable that in recent work on a similar phenomenon in Chinese — the grammaticalisation of *de* ‘get, obtain’ — Sun Chao Fen (1996) suggests that the first step in the grammaticalisation of the main verb *de* ‘get, obtain’ was its appearance in V1 position, i.e. as a complement-taking predicate.) (Note also that the scenarios in Figures 3 and 4, above, are oversimplified, since postverbal *baan* as we know it has at least two separate meanings, raising the question as to which one of these meanings — i.e. ‘can’ or ‘succeed’ — is the source of the putative extension from post- to pre-verbal *baan*.)

Once we entertain the possibility that the hypotheses illustrated in Figures 4 and 5 may be valid, we call into question the very premise of Haiman’s paper, i.e. the need to explain a strange/anomalous ‘migration’ of a morpheme from postverbal position to preverbal position. Let us thus consider the likelihood that preverbal (V1) *baan* developed directly from a main verb. In explaining the initial emergence of *baan* in V2 position (i.e. Step 2 in Figures 2–5, above), Haiman appeals to Khmer ‘verb-concatenation’ as an enabling factor. The same argument may also be used to explain the appearance of a main verb *baan* in V1 position — as a complement-taking predicate — without having to posit a strange process of ‘migration’. Haiman does not say why he rejects the simpler main-V *baan* > pre-V *baan* hypothesis (Figures 4–5, above), and yet the suggestion of ‘migration’ is far-fetched in comparison to the simpler and more familiar process of a transitive verb extending its function to accept verbal complements as well as nominal ones (i.e. a transitive verb becoming a complement-taking predicate as well). This process may have been followed by reanalysis of the complement-taking predicate (i.e. V1 as clausal head in a V1-V2 string) as a preverbal marker (i.e. V2 becomes clausal head, and V1 a pre-verbal marker), a straightforward process in a language like Khmer in which clausal subordination is not necessarily given overt morphological marking (e.g. by a non-finite verb form in the subordinate clause). Indeed, Haiman’s important point about initial position attracting phonological stress/attrition in Khmer and similar strongly head-initial neighbouring languages may be put to better use in explaining how erstwhile complement-taking predicates so often become preverbal aspect/modality markers in languages of the region (cf. the deverbal ‘auxiliaries’ Haiman mentions on p. 154).

The first serious problem with Haiman’s contribution is this unquestioned assumption of a strange historical process which needs explaining. If we are to seriously consider Haiman’s concluding speculations relating to universals of typology and grammatical change, then it must first be *established*, not merely *assumed*, that the ‘migration’ his paper seeks to explain (and from which he

draws theoretical conclusions), actually happened. It is unlikely that any such ‘migration’ ever took place.

A second important problem with Haiman’s paper concerns the standard of semantic/grammatical description. A prerequisite for speculation about diachronic semantic/grammatical processes is precise and well-supported synchronic description which clearly and explicitly separates the *semantics* of the relevant forms from their *pragmatically enriched import*, in given contexts. This distinction is indispensable in studies of grammaticalisation since it is the very interplay between effectively ‘fixed’ semantics and predictable inferences that gives rise over time to new and separate meanings, and to associated changes in grammatical structure. However, the semantic description in Haiman’s paper is neither adequately clear nor adequately supported by the data and arguments provided. First, in describing various meanings of *baan*, Haiman does not distinguish between invariant meanings (entailments, effectively stable across contexts) and context-dependent interpretations (common inferences, context-dependent translational equivalents). The result is persistent unclarity as to whether a given ‘meaning’ of *baan* is actually an entailment of the lexical item in Khmer, or a contextual interpretation drawn from the English translation provided. A second problem is that the overall description of *baan* and its functions is too coarse, failing to mention a number of distinct semantic and grammatical functions of *baan*. Rather than commenting in detail on the way Haiman treats his data, I refer the reader to my own description of the polyfunctionality of *baan* in Khmer, among similar morphemes in other languages of the region (Enfield 2000). It is necessary in this context, however, to point out a few of the more significant problems.

As a main verb, *baan* is misleadingly characterised using ‘seize’ and ‘catch’ (among better translations), both of which are inappropriate due to the specifically non-agentive/non-controlled nature of main verb *baan*. *Baan* as a main verb ‘get, acquire’ cannot be used in imperative expressions, and cannot appear with adverbs of volition/control, such as ‘carefully’. As a main verb, *baan* predicates something happening, not someone doing something. This is important for Haiman’s later claim that ‘auxiliary’ *baan* has a ‘causative’ function.

Post-verbal *baan* is described as a ‘success verb’ in resultative compounds’ (p. 149), and Haiman insists on calling it a ‘success verb’ throughout the paper. However, we see many [V *baan*] expressions in which there are not *two events* referred to at all (e.g. some goal-directed action in V1, with V2 predicating resultant/consequent success), and in which ‘success’ is not even predicated anyway. Postverbal *baan* often means ‘can’ and nothing more. Thus, in all but

two of the ten examples Haiman gives in his (2) and (3) (p.152–3), we can attribute to postverbal *baan* the invariant meaning of ‘can/able to’, with the ‘manage to’/‘succeed in’ meanings in Haiman’s translations arising pragmatically in context. Consider, for instance, Haiman’s translation of (2c) — ‘If you don’t succeed in ordering him around...’ — which is coloured with context-based interpretation, beyond the basic meaning of ‘If you are not *able to* order him around’. Examples (2a, d) exemplify a second meaning of postverbal *baan*, i.e. ‘V1 with success’. Haiman’s (2d) nicely illustrates the polysemy of postverbal *baan*, but this is obscured by the misleading gloss of *sdap* as ‘understand’. *Sdap* means ‘hear’, and the combination ‘hear’+*baan* has two meanings (due to the polysemy of postverbal *baan*). It can mean either ‘can hear’ (i.e. the sound is audible), or ‘successfully hear’ (i.e. ‘can understand (what someone says)’). Importantly, *sdap baan* [hear “get”] cannot be used to predicate ‘understanding’ in contexts where the subject does not *hear* the thing s/he is attempting to understand (e.g. when unable to understand something s/he is *reading*). (The second example concerns the combination ‘seek’+*baan*, which may mean either ‘can seek’, or ‘successfully seek’ (i.e. ‘find’).) Haiman further misleads by declaring that postverbal *baan* ‘can’ is ‘necessarily invisible’ (p.154) when it occurs in combination with the preverbal element *qaac* (which also means ‘can’). But this claim makes no sense, since in real language dual assertion of a single meaning in a single proposition does *not* amount to communicative vacuity. On the contrary, there is an important pragmatic function of this double marking of ‘can’; it creates a more wordy and more ‘precise’ sounding expression, which can achieve both (a) emphasis on expression of the meaning ‘can’ (often with the result of defeating/precluding the common inference of ‘success’), and (b) a formal turn of phrase (Haiman’s (3d,e), p.153, are good examples).

Haiman’s description of preverbal *baan* suffers from similar problems. The reader is given a string of examples of *baan* as a preverbal ‘auxiliary’, with ‘at least four, possibly five’ different meanings (p.155–8). But these ‘different meanings’ are not established by the examples Haiman gives, rather they are drawn from his free English translations. I suggest that one robust invariant meaning of *baan* V (covering two of Haiman’s ‘meanings’) is approximately

*V; this is the case because something else happened before this*

which definition provides a single meaning covering all of the generously enriched ‘have a chance to V’, ‘get to V’, ‘chance to V’, ‘have the fortune to V’, ‘come to V’, ‘become V’ expressions in Haiman’s free English translations.

A second function of pre-V *baan* — Haiman’s ‘past tense/already’ function — is to give emphasis to the ‘reality’ of a predicate (i.e. as an emphatic/realis marker), like English *do* in an expression like *He did go* (as opposed to *He went*). A significant point that Haiman does not note is that preverbal *baan* in this function is much more common in formal contexts, such as the kinds of traditional tales his 2.3.2 (p. 156–6) examples are from, or in newspaper reports, or other official/‘high’ contexts. This accords with preverbal *baan* being emphatic (sounds more precise), and wordy (sounds more formal/polite). Where Haiman’s analysis of the Khmer data becomes too flimsy is in his 2.3.4 (p. 158), where he suggests that the preverbal ‘auxiliary’ *baan* has a ‘causative’ function (‘get’/‘cause’/‘make’). In none of his (9) (p. 158) examples does *baan* itself perform a ‘causative’ function in any usual sense of that term (e.g. in which the morpheme facilitates addition of a new argument to the clause; in which some coercion/enablement is predicated; etc.), and most importantly, in none of the examples is *baan* even adjacent to a main verb!

A serious problem in Haiman’s treatment is what he does *not* say about *baan*, i.e. his failure to mention a number of distinct functions *baan* performs. This is made more difficult to ignore by the fact that some of the examples Haiman provides exemplify these other uses of *baan*. Consider (5a) (p. 154, Haiman’s numbering/transcription/glosses/translation):

- (5) a. *koon krɔbɔj dɛɛl tɔp kaet baan dɔp tɕaj* ...  
 baby buffalo that just born ? ten day  
 ‘A baby water buffalo just ten days old’ (*baan* = ‘past tense?’ ‘ago?’)

Haiman is unable to attribute a function to *baan* in this example, because he insists on treating it as an instance of post-V *baan*. But while *baan* does appear after a verb, this is not a mere ‘postverbal’ occurrence. (The whole phrase *baan dɔp tɕaj* [get ten day] is omissible without changing the basic semantics of the main clause, while *dɔp tɕaj* [ten day] is not.) *baan dɔp tɕaj* [get ten day] in this example is a post-verbphrasal adverbial adjunct in which *baan* introduces a period of time since some state/event, predicated by the main clause, has become the case. A more faithful translation of (5a) would be ‘a baby water buffalo which was just born (since) ten days ago’. This is a distinct and common function of *baan*, one which does not grammatically behave like a postverbal modal.

Another, similar, function of *baan* is to head postverbal adverbial adjuncts of manner, as in Haiman’s (9b) (p. 158), [fold blanket keep put (*baan* good)], in which ‘*baan* good’ means ‘well’ or ‘nicely’. Haiman strangely interprets this as a preverbal ‘causative’ function of *baan*, although *baan* in this example

neither marks a main/core verb, nor introduces an argument into the clause (which functions one would expect a ‘causative’ morpheme to perform). (Indeed, *baan* never has a valency-increasing function, weakening Haiman’s claim that *baan* may have a ‘causative’ function, and may be related to the Khmer morphological causative prefix *bVN-*.)

A final oversight is Haiman’s failure to make a distinction between (a) the preverbal *baan* which appears *immediately adjacent to the main verb* (i.e. at an inner layer of the clause), and (b) the *baan* which appears *before the subject of the clause* (i.e. at an outer layer of the clause), the latter often linking to a previous clause, in subordination (as in (6b), p. 155), and often in combination with *ciə* ‘be’, as many of Haiman’s examples show. Sentences (6b), p. 155, and (9a,c), p. 158, are good examples of *baan* operating over not just the verb, but the whole following clause. According to my own investigations (Enfield 2000), the semantic import of this pre-subject outer-layer/clause-linking *baan* is quite similar to that of one preverbal function of *baan*. Even so, the pre-V and pre-subj functions of *baan* are different, and must be recognised as such. Especially relevant to Haiman’s interest in preverbal *baan* becoming fused to the verb and there becoming phonologically reduced is the fact that in one function (i.e. inner-layer, pre-V), *baan* is immediately adjacent to the verb, while in another (i.e. outer-layer, pre-subj), it is not. This is a crucial point which ought to surface in Haiman’s discussion.

I will only briefly address Haiman’s novel suggestion that a predecessor of ‘causative’ preverbal *baan* may have been the source of the Khmer *bVN-*causative prefix. This hypothesis, as Haiman presents it, depends in the first place on *baan* having a ‘causative’ function (in particular, where a new argument is added to the clause, as in the many examples of valency-increasing derived *bVN-* forms on p. 159–60), and more specifically, having this function in a position immediately adjacent to the main verb of the clause (such that it may fuse to it and there phonologically reduce). As noted above, Haiman’s supposed ‘causative’ examples of pre-V *baan* in (9) (p. 158) do not show *baan* performing a valency-increasing function (nor are they ‘causative’ in the sense of predicating coercion or enablement with respect to the event or action predicated by the marked verb), and further, in none of the three examples he gives does *baan* actually occur adjacent to the main verb of the clause. This weakens the hypothesis so much that it cannot be taken seriously without further support. In addition, we are faced again with the error of temporal logic described above, namely that contemporary forms are being regarded as historical sources for contemporary forms. All we are left with is phonological



similarity between *baan* and the *bVN*-causative prefix, and little else of substance. Haiman's appeals to the entirely vague concept of 'desemanticization' do not help. Anyway, withholding any judgement as to whether Haiman's hypothesis of *baan/bVN*-relatedness could be correct or not, a couple of further points should in the meantime be included in Haiman's discussion. First, recall that *baan* as a main verb is specifically *non-controlled/non-agentive* — it may not appear as main verb in an imperative expression, and may not be marked by controlling adverbs like 'carefully'. Recalling also that preverbal *baan* may have developed directly from a predecessor of this non-agentive/non-controlled main verb, this makes it an unlikely source for the kind of agentive/controlled causativity found in *bVN*-derived expressions like 'teach', 'separate', 'dilute', and 'loosen' (see Haiman's Section 2.4, p. 158ff). Second, while preverbal *baan* expressions glossed as 'get to V', 'chance to V', etc., do include 'because' in their semantic structure (see my proposed definition above), the structure of 'causation' (if that term is applicable at all) is quite different to that of *bVN*-causative derivations. In an expression of the form [NP *baan* V] (disregarding an emphasis/realis/past/'already' reading), the whole predication NP V is *enabled by* or *caused by* some (understood) prior event/situation, and *grammatical relations are not altered* — NP remains the agent of V, with no new argument added to the clause. But in an expression of the form [NP *bVN*-V], NP is *not* the agent of V at all, and it is NP's action that *causes* or *enables* the action or event predicated by V (with a causee/lower agent of this V added to the clause, and expressible in object position; see the examples supplied by Haiman on p. 159–60).

In submitting this criticism of Haiman's work, my purpose is not to dispute his general conclusions or theoretical speculations, but I do contend that his arguments are based at least in part on assumptions which are inadequately justified, and that his conclusions do not follow from the data and argumentation supplied. First, the post-V *baan* > pre-V *baan* 'migration' hypothesis — the very premise of the paper — is unquestioningly adopted in favour of at least one simpler and more natural scenario, namely that preverbal 'auxiliary' *baan* developed from a complement-taking predicate, which, prior to that, developed directly from a main verb 'acquire'. Other scenarios are possible. Second, semantic relations which are first argued to hold between forms *synchronically* are then unquestioningly assumed to hold *diachronically*, a leap unwarranted without supporting argument. Finally, before we can even consider the validity of Haiman's speculations, it is necessary to first properly establish the functions of *baan* with careful and well justified description of both the semantics and pragmatics of this noteworthy polyfunctional item.

## References

- N.J. Enfield, 2000. *On the polyfunctionality of 'acquire' in mainland Southeast Asia: a case study in linguistic epidemiology*. University of Melbourne dissertation.
- N.J. Enfield, in press. On genetic and areal linguistics in Mainland Southeast Asia: parallel polyfunctionality of 'acquire'. In *Genetic and Areal Linguistics*, edited by Alexandra Aikhenvald and R. M. W. Dixon, Oxford: Oxford University Press.
- Sun Chao Fen. 1996. *Word-order change and grammaticalisation in the history of Chinese*. Stanford University Press.

### *Author's address*

N.J. Enfield  
Language and Cognition Group  
Max Planck Institute for Psycholinguistics  
PB 310 AH-6500 Nijmegen, The Netherlands  
Nick.Enfield@mpi.nl