

## Guidelines for contributions to the Anglicism database (GLAD)

### General

All data should be submitted as an Excel file based on the official GLAD template

**GLAD-Anglicisms-template-after-Caceres.xlsx** to [glad@nhh.no](mailto:glad@nhh.no), taking heed of all points mentioned in these Guidelines.

### Lexical selection

- a) In the construction phase of the database, Anglicisms should be selected according to the initial letter of the **etymon**, and deliverables should target all Anglicisms whose etymon start with a particular letter (e.g. letter A-C in spring 2018, etc.). This means that, for example, in several European languages, certain Anglicisms starting with the letter K should be treated as letter-C items. A case in point is Slv. *kompjuter* 'computer'. Similarly, loan translations like Nor. *nedlaste / laste ned* 'download' are letter-D items. Contributors should aim at compiling the critical mass of Anglicisms for each letter in their respective language.
- b) **Types of Anglicisms** to be included in the GLAD database are the following:
- i. *Unadapted borrowings*, including multi-word units, acronyms, terms originating in non-English speech communities in the Anglosphere (e.g. *tomahawk* and *jungle*), and internationalisms coined in English
  - ii. *Adapted borrowings*, representing the same types as above
  - iii. *Semantic loans* (domestic words taking on English senses)
  - iv. *Loan translations*, including phraseologisms (unit-for-unit translations of English compounds/multi-word units)
  - v. *Hybrids*, including productive affixes (domestic compounds with at least one English component)
  - vi. *Pseudo-Anglicisms*: resemantization (domestic sense for English word, e.g. Ger. *handy*), clippings (shortened English words, e.g. Dan. *happy end*), and recombinations (domestic combinations of English elements; e.g. Jap. *akuhara* 'alcohol harassment')
  - vii. *Phono-semantic matchings*, where an English lexical item is matched with a phonetically and semantically similar pre-existent domestic word (cf. Ger. *Was gibt's?* → Eng. *What gives?*)
- c) **Not to be included:**
- i. Proper nouns and brand names
  - ii. Name-based derivations (e.g. domestic adjectives based on English proper nouns, like *Orwellian*)
  - iii. Frequency-boosted domestic words (their increased usage due to similarity with English etymons)
  - iv. Obsolete items (see criteria under point d)
  - v. Items from a non-anglophone speech community mediated via English (e.g. *sushi*, *nachos*)
  - vi. Specialist vocabulary not used in the general language.
- d) **Types of discourse** serving as a basis for inclusion of items:  
**General** language, whether spoken or written, including common slang and determinologized items, but excluding in-crowd jargon, solely technical lingo, and dialect-specific features.

**Modern** language as written or spoken in the 20<sup>th</sup> and 21<sup>st</sup> centuries. This means that even Anglicisms attested before 1900 could be included if they were – or still are – commonly used after 1900.

## Data categories and submission

- a) **The official GLAD template** (presently post-Caceres revision) should be used consistently. This is downloadable from the Resources page on the GLAD website ([www.gladnetwork.org](http://www.gladnetwork.org)). To ensure highly informative and comparable content, make sure that for each item, all **five obligatory cells** are filled in: Language, Domestic form, Etymon, Type of borrowing and Part of speech (POS). Feel free to fill out as many of the fourteen optional cells as you see fit. For several categories, the template has **comment fields** that emerge when you hover the mouse above the category label. All such comments should be consulted before you start filling in the template. When you enter data in the template, some cells ask you to choose one of a set of **pre-defined labels**. Thus, when you place the cursor in a cell in the 'Type of borrowing' column, a **small arrow to the right of the cell** lets you choose between the seven types listed under point b) above. The categories with such pre-defined labels are Type of borrowing, POS, Phrasemic type and Gender.
- b) The word chosen as **Domestic form** should be a standardized and commonly accepted variant of the Anglicism in the recipient language. Only one form should appear in this category. If there are other variants, these should be listed under Domestic form variants and separated by a comma + space.
- c) The **Etymon** field contains the English etymon with the orthography which the OED lists as its head form. It is crucial that etymons are harmonized across languages for linking purposes within the database. OED uses the letter z for Greek-based words with suffix *-ize/-ization* etc., so GLAD will have the z-form as its etymon in words like *harmonize*. Furthermore, hyphenization depends on OED's standard, so a word like *babysitter* should not have a hyphen while *anti-American*, *co-operation* and *thought-provoking* should. NB! For recombinations or multi-word pseudo-Anglicisms which have no English correspondent, **the symbol + may be used** in the Etymon field to indicate that this is not an originally English word or collocation; e.g. Nor. *snacksy* – *snacks+y*; Jap. *aruhara* – *alcohol+harassment*. For clipped pseudo-Anglicisms we use the full English etymon even if the domestic form is clipped, e.g. Dan. *butterfly* – *butterfly bow-tie*. Under normal circumstances, grammatical words like the definite article or infinitive marker are left out of etymons (exception: the expression *to be or not to be*, where the infinitive marker is an integral part).
- d) **Definitions** are semi-obligatory, that is, a definition must be included for words with more than one meaning (polysemy). In order to save time and to make definitions comparable across languages, it is recommended to use the latest edition of the Concise Oxford Dictionary as a common point of reference and take over the paraphrases for the various senses given there.
- e) GLAD uses a relatively narrow set of **POS tags**, with noun, verb, adjective, adverb, interjection and other as the only possible values. The determining factor for POS is the word's grammatical function in the recipient language, i.e. Ger. *handy* = noun. For phraseological units, it can be difficult to decide, and many of them will be classified as other. Note that ordinary multi-word compounds such as *butterfly effect* and *Black Friday* are not technically treated as phraseological units in GLAD; they are simply noun+noun compounds, i.e. they should have the POS label noun and not be assigned a Phrasemic type in column K;

cf. f) below. The same goes for so-called irreversible binominals, like *bed and breakfast*, which to all intents and purposes are nouns. Similarly, phrasal verbs are treated straightforwardly as verbs, e.g. Nor. *føkke opp* 'fuck up'.

- f) The category **Phrasemic type** is to be used for various types of phraseological units, but not for phrases that function straightforwardly as nouns, verbs etc. as described in e). The rather coarse classification that we use is a simplification of the taxonomy proposed by Granger & Paquot (2008). Below is a list of the labels we use.

G&P label	GLAD label	Definition	Example
Idioms	idiom	constructed around a verbal nucleus	<i>to spill the beans</i>
Similes	simile	stereotyped comparisons	<i>fit as a fiddle</i>
Textual sentence stems	discourse marker	routinised fragments with textual function	<i>the thing is that; it will be shown that</i>
Speech act formulae	discourse marker	routine formulae with discourse-pragmatic funct.	<i>good morning; take care; how do you do</i>
Attitudinal formulae	discourse marker	signal speaker attitude	<i>in fact; to be honest</i>
-	discourse marker	expletives	<i>What the fuck!</i>
Commonplaces	commonplace	non-metaphorical sentences, truisms/ tautologies	<i>Enough is enough; YOLO</i>
Proverbs	proverb	express general ideas non-literally	<i>When in Rome ... ; The early bird ...</i>
Slogans	slogan	directive phrases used repeatedly in politics/ advertising	<i>Make love, not war; 'I'm lovin' it'</i>

For many of these, the correct POS label will be other, unless a more specific label can be assigned, e.g. *fit as a fiddle* – adjective; *spill the beans* – verb.

- g) **Pronunciation** is a semi-obligatory field; it must be included in cases where the pronunciation of an Anglicism deviates considerably from the phonotax of the recipient language. The pronunciation should be given in IPA and it must show how a given word is pronounced in the recipient language.
- h) **Frequency** should be indicated thus: \* = quite rare; \*\* = frequent; \*\*\* = highly frequent. Frequencies for languages without major accessible corpora will have to be partly based on compilers' personal judgment.
- i) **First attestation** should be indicated by means of the most appropriate 4-digit year. Approximations are accepted. For reasons of cross-comparability we disallow Görlach-style dates such as '20c', '1980s', 'end19c', so these should be converted to the acceptable GLAD format, e.g. end19c → 1890.