The Grammaticalization Cycle

Languages fall onto a spectrum according to

- how much of their grammar is **morphology** (i.e., endings, paradigms, inflections, agreement). These with more are called **inflected** or **synthetic** languages; Latin and Eskimo are examples, the latter an extreme example called **polysynthetic**, where the distinction between word and sentence is weak.

- and

- how much of their grammar is **syntax** (i.e., word order, constructions, particles, prepositions, idioms). These with more are called **isolating** or **analytic** languages; English and Chinese are examples.

Typically the more a language has of one kind of grammar, the less it has of the other. There’s an optimal balance between convenience and efficiency, which overlap considerably; this balance is always shifting, like balance in any living system. Old English had 4 noun and pronoun suffix cases comprising various paradigms that were applied to just about every noun and pronoun, and used daily in every sentence.

1000 years after the Middle English period began, fewer than a dozen fossils of the Old English paradigms are left in Modern English:

- the **genitive** ‘apostrophe-S’ suffix, which has fossilized into a clitic -Z₂, identical in form to the noun plural suffix -Z₁, but attaching to a noun phrase (the King of England’s wife) rather than to a noun (*the King’s of England wife), the way a real case would work.

- the **dative/accusative** suffix, now often called ‘objective’, which now is a living fossil, occurring on only **five** words, all pronouns, that are special objective forms:
  - *he* ~ *him*
  - *we* ~ *us*
  - *they* ~ *them*
  - *I* ~ *me*
  - (and sometimes) *who* ~ *whom*,

which has been dead for ages and is a zombie word now.

All Proto-Indo-European inflection was suffixal. PIE and all of its daughter languages (Sanskrit, Greek, Latin, Proto-Germanic, etc) were synthetic and heavily inflected – Sanskrit had 8 suffixal cases and 3 voices, for instance. English, however, has lost almost all its inflections, and is now a textbook analytic language. How did this happen? The real answer is that nobody knows; it’s too complicated and there’s too little data. There are, however, theories. Here’s one, called **Grammaticalization**, that clearly explains some things.

Most affixes occur at the ends of words; this is also where most sound changes happen. When a phonetic change occurs at the end of a word, like final –m and –s going silent in Vulgar Latin, a lot of the paradigmatic endings that Latin inherited from PIE disappeared, leading to the modern Romance languages. This moves Latin (catastrophically) from being a synthetic (inflected) type language to being an analytic (uninflected) type.

The process is called **grammaticalization**, and refers to the gradual loss of affixal inflections to sound change, followed rapidly by new syntactic constructions with auxiliary words to fill the gap, followed in time by new inflections based on common constructions with auxiliaries.

The example below happened a thousand years before Old English changed into Middle English, but it’s similar.

Classical Latin had 5 cases for most nouns, distinguished by suffixes, thus:
coulda, hadna, usta, hafta, lookit, etc.

Language becomes analytic.

Idioms and syntactic constructions become fixed and abbreviated.

At, in this case, is probably 1 kiloyear or so, as Old English → Modern English shows. The bottom arrow is what has started to happen in French (which is becoming a prefixal language), and in the English eye dialect words gotta, wanna, oughta, shoulda, shouldna, coulda, hadna, usta, hafta, lookit, etc.