Regular and Irregular Forms

Why do languages have irregular forms?

Irregularity in language can have a number of causes:

- It can be a historical remnant (e.g. ***child*** – ***children***, left over from Anglo-Saxon).
- It can be a recent borrowing (e.g. ***couture*** – ***couturier*** from modern French).
- It can be a grammatical quirk (e.g. non-count nouns like ***sand*** can be both singular and plural).
- It can be the result of two different forms merging into a single word complex (e.g. ***go***: from Saxon ***gan***, to walk; ***went***: from Anglish ***wenden***, to proceed).
- It can be a marker of cultural usage (e.g. some American dialects retain the old ***got*** – ***gotten*** form, while most English dialects use ***got*** – ***got***). This is sometimes known as the **shibboleth effect** (see the Bible, Book of Judges, chapter 12 for the story).

Where does irregularity happen?

Irregularity happens with common words and rare words, but for different reasons. There are sometimes irregularities in words which fall between common and rare, but they tend to become regular over time. For instance, the Anglo-Saxon past tense ending of *–t* is being slowly replaced by the regular *–ed*. ***darnt*** and ***spurnt*** have already become ***darned*** and ***spurned***, and ***burnt*** is in the process of becoming ***burned***; but ***slept*** has not become ***sleeped*** (possibly because of the vowel mutation) and ***spent*** has not become ***spended*** in the UK, although it has in parts of the USA.

Irregularity in uncommon forms is usually maintained when they are still recognised as borrowings. So the plural of ***formula*** is ***formulae***, because we recognise its Latin origin (although ***formulas*** has become an acceptable alternative). Similarly, ***biscotti*** is an acceptable plural and, in this case, ***biscotts*** is not. However, a word usually regularises if it ceases to be commonly known as a borrowing, (e.g. ***thug*** – ***thugs***, not ***thug*** – ***thuggee***).

Irregularity in common forms is usually maintained by the fact they are regularly used, and aberrations from the norm are easily recognised and remembered. Some common verbs are therefore irregular in many languages, including ***be, go, come, have, give, do, make, want, need, must, say, see, know, think, tell, choose, stand*** and ***sit***.

Common personal nouns can also have irregular forms, such as ***man, woman, child***. Less reliably, irregularity can cluster in other classes of nouns: body parts (e.g. ***foot, tooth***); common animals (e.g. ***goose***, ***mouse***, ***sheep, ox***); and common food items (e.g. ***loaf***).

One cluster of irregularity is in pronouns. Most languages have a pragmatic approach to pronoun formation, and there is quite a large variation in the range of pronouns. For instance, English has five pronouns, with the singular and plural second person being the same word; French second person divides into singular informal (***tu***) and singular formal plus plural (***vous***); Spanish has a whole extra case for the second person formal (***usted/ustedes***); and BSL has at least three different signs for first person plural (me and you, me and someone else not you, and me and a group). The first person plural is also used pragmatically in many languages as a formal version of self-reference, or to indicate an undefined group which may or may not include the speaker.

Does your language need to be irregular?

There is no rule that says a language will always have irregularity built in – some natural languages are quite regular. For instance, Afrikaans has several sources (Dutch, English, Arabic, Isizulu, Sotho, Ndebele and Swazi, among others). However, where it has borrowed words, it has applied the same simple grammaticalization to them, making it a very regular language. Similarly, Haitian Creole has abandoned many of the irregularities of French and its other source languages. Of course, created natural languages like Esperanto are probably the most regular – but, with usage, even they begin to develop idiomatic forms.

A reasonable rule is that new languages tend to be more regular than old ones, but languages which continually reinvent themselves (like English) also tend to continually regularise themselves. No language can be more irregular than the human mind can handle, so that imposes a natural limit on complexity; but nobody has tested where that limit lies.

Why Grammar Isn't Everything

Chomsky has proposed that the central component in linguistics is grammar:

> The phonological component of a grammar determines the phonetic form of a sentence generated by the syntactic rules. That is, it relates a structure generated by the syntactic component to a phonetically represented signal. The semantic component determines the semantic interpretation of a sentence. That is, it relates a structure generated by the syntactic component to a certain semantic representation. Both the phonological and semantic components are therefore purely interpretive. Each utilizes information provided by the syntactic component concerning formatives, their inherent properties, and their interrelations in a given sentence. Consequently, the syntactic component of a grammar must specify, for each sentence, a *deep structure* that determines its semantic interpretation and a *surface structure* that determines its phonetic interpretation. The first of these is interpreted by the semantic component; the second, by the phonological component.
>
> Noam Chomsky (1965). *Aspects of the Theory of Syntax*. MIT Press: Cambridge, MA, USA, p16.

However, there are aspects of semantic interpretation which do not seem to rely on structure at all, they rely on context. This context can be cultural, lexical, phonological, or systemic, as the following examples show.

CULTURALLY CONTEXTUAL MEANING: I gave the man his greetings

We know instinctively that *the man* is different from the person who generated the greetings because, in English culture, people don't greet themselves. So we can accept the grammatical reordering of **I gave his greetings to the man** as having the same meaning. However, the construct, **I gave the dog his dinner** is polysemic: *his* usually refers to the dog, because the act of giving creates ownership by the dog; but it can refer to another person. However, the construct **I gave his dinner to the dog** has only one meaning, because the adpositional phrase distances *the dog* from *his dinner*, making *his* refer to another person.

LEXICALLY CONTEXTUAL MEANING: Alf saw a house with …

(This was discussed in lecture four.) The grammar of "with" in this construct cannot be resolved until the semantics of the following noun phrase are resolved. Lexis can carry semantic information and determine grammar.

- **Alf saw a house with chimneys** – the house has chimneys, so "with" links *chimneys* and *house*.
- **Alf saw a house with Beth** – Beth is a person, so "with" links *Beth* and *Alf*.
- **Alf saw a house with binoculars** – binoculars are tools for seeing, so "with" links *binoculars* and *saw*.
- **Alf saw a house with surprise** – the surprise was the outcome of Alf seeing the house, so "with" links *surprise* and *Alf saw a house*.

PHONOLOGICALLY CONTEXTUAL MEANING: Asking a question

An English question can be formed by grammatical movement: **Alf is a linguist** becomes **What is Alf?** by the conversion of the object word to a WH-word and the transposition of subject and object positions. However, it is perfectly reasonable in English to use **Alf is what?**, marking the question form with a rising inflection rather than grammatical movement. Similarly, asking **is Alf a linguist?** with a rising tone on the last syllable implies the question is about new information; while **is Alf a linguist?** with a falling final tone implies the question is confirming information. Phonology can therefore carry semantic information independent of the grammar.

SYSTEMICALLY CONTEXTUAL MEANING: Alf and Beth are Cornish and Welsh

This construct has a logical form of **[(A + B) = (X + Y)]**, but we know that being X usually precludes being Y. So the construct has a culturally bound meaning that **[A=X and B=Y]** but this meaning is also logically open to the meaning that **[A=Y and B=X]**. To test this, people use the construct **which is which?** "Which" has a semantic role of a choice between two alternatives, so the logical meaning of the construct is **[(A or B) = (X or Y)?]**, but its actual meaning is **[(A=X and B=Y) or (A=Y and B=X)?]**. Our cognitive logic system can determine meaning independently of grammar.

PULLUM'S PASSIVES: simple movement doesn't always work

The passive is formed by movement, as in **Alf gave a lecture → A lecture was given by Alf**. But not always:

- **Mike seemed a nice enough guy → *A nice enough guy was seemed by Mike;**
- **The fish weighed twelve pounds → *Twelve pounds were weighed by the fish**;
- **This matters a lot to me → *A lot is mattered by this to me**.

Geoffrey K. Pullum (2007). Creation myths of generative grammar and the mathematics of Syntactic Structures. In Christian Ebert, Gerhard Jäger & Jens Michaelis (eds.), *The Mathematics of Language*. Springer: Berlin, Germany.