Inflection, Derivation, and Compounding

David R. Mortensen

January 27, 2025

Introduction

The prototypical morphological operation adds some formal material (some phonemes) to a base and, in doing so, creates a new word with a different meaning:

(1) a. drink
b. drinker

(2) a. truck
b. trucker

The kind of operation is called DERIVATION. But other morphology is different. In some cases, the operation does not create a new word (but only creates a form of the word that fits a particular grammatical context). For example, the word DOG may be either *dog* (singular) or *dogs* (plural). This is called INFLECTION. In still other cases, rather than adding an affix to a base, a new word is formed by concatenating two words or stems, as in *catfish* or *dishwasher*. This is called COMPOUNDING. Derivation and compounding, together, are called word-formation.

Stems, as we will learn, are roots plus zero or more derivational affixes.

It is useful to understand these three categories, and why they have been so important historically. It is also important to know that they exist on continua.

The distinction between derivation and inflection

A prototypical derivational operation behaves rather like a function that operates on a graph (or some other mathematical representation), where the graph represents the meaning and part-of-speech of the word. The domain of this function could vary greatly (the operation cannot apply to all bases) and so can the change brought about by the operation. Consider the suffix *-ism*: it can apply only to certain kinds of adjectives and nouns (*social*, *capital*, *federal*, *Lenin*, etc.). It has a very specific meaning (to the extent that we expect most languages will not have equivalent morphology).

Contrast this with a prototypical inflectional operation, which acts like a function that flips a bit in a value associated with the word. There are a finite number of bits in the inflectional value, each corresponding to one of the finite sets of properties like case, gender, number, voice, tense, aspect, and modality [1] that can easily be expressed as "on" or "off" (or as multi-valued features).

[1] We'll illustrate these later in the lecture

Table 1 lists some differences between prototypical derivation and prototypical inflection. A similar table is found in Haspelmath and Sims[2]. In general, a prototypical derivational operation changes meaning and part of

[2] Martin Haspelmath and Andrea Sims. *Understanding Morphology*. Hodder Education, London, 2nd edition, 2010

PROTOTYPICAL DERIVATION	PROTOTYPICAL INFLECTION
changes concept	does not change concept
changes part of speech	does not change part of speech
not directly relevant to syntax	directly relevant to syntax
extrinsically ordered	intrinsically ordered
non-paradigmatic	paradigmatic

Table 1: Prototypical derivation versus prototypical inflection

speech in a way that is transparent to the syntax. That is to say, a words that becomes a noun through one derivational processes (e.g., adding *-ness*) behaves the same way in the broader grammar as a word that becomes a noun through another derivational process (e.g., additing *-ity*). The order of a prototypical derivational operation is based on principles of scope (just like mathematical operations) and different ordering of operations can result in different meanings but the order of prototypical inflectional operations is fixed and follows from general principles. Finally, proto-typical derivation does not produce PARADIGMS. In the sections on derivation and inflection, we will look at these differences in more detail.

Paradigms are multidimensional arrays of word forms where each row or column corresponds to a particular value for a particular feature. A feature, in this sense, can be something like NUMBER, which might take values—as in Old English—of SINGULAR, DUAL, and PLURAL.

The distinction between derivation and compounding

Derivation differs from compounding solely in that compounding combines words/stems while derivation involves an operation on a base (like suffixation). This may seem very simple. However, derivation often arises historically from compounding.

Compounding

Endocentric versus exocentric compounds

An important term when speaking of compounds is HEAD. Heads determine the basic semantics and part of speech of a compound. For example, *blackboard* is a noun like *board* (not an adjective like *black*) and is a kind of board, not a kind of black. Therefore, we say that *board* is the head of BLACKBOARD and that this word is head-final. One way of verifying the head of a compound is via implication: if being a blackboard entails being a board then *board* is the head of *blackboard*. Likewise, being a bathroom entails being a room (but not a bath), so we say that the head of *bathroom* is *room*. This test is not definitive in the case of *blackboard*, though, since being a blackboard also entails being black. The issue is settled, in that case, by the part-of-speech criterion.

In most compounds, one of the words is the head (*room* or *board*). These are called ENDOCENTRIC compounds. Examples from include *cable car*,

gunmetal, *wallpaper*, and *football*. However, there are other compounds where neither of the words is a head. It is as if the head sits outside of the compound. Examples include *greybeard*, *redhead*, and *spendthrift*. Consider, also, these famous examples from Italian:

Note that English compound words are sometimes written with no delimiter (as in *football*), sometimes with hyphens (as in *single-minded*) and sometimes with a space between the components (as in *ice cream*. This is an orthographic difference rather than a grammatical one.

(3) a. porta + bagagli
carry + luggage
'trunk'

b. lava + piatti
wash + dishes
'dishwasher'

c. asciuga + capelli
dry + hairs
'hair dryer'

The signified, in these compounds, is predictable from the signifier. However, they are not exactly compositional. Rather, they involve a construction that combines a transitive verb V and a noun N that serves as the semantic PATIENT of the verb. The meaning of the whole is an kind of object that V's Ns. They are, expressly, not kinds of Ns or kinds of Ving.

transitive verb is one that takes an object as well as a subject. *Die* is an intransitive verb and *kill* is a transitive verb

A patient is a participant in an event to which or to whom the event happens.

Exocentric compounds are good examples of non-compositional constructions.

Subordinate versus coordinate compounds

A compound may have a single semantic and syntactic head (an endocentric or *tatpuruṣa* compound). It may also not have a semantic head (an exocentric or *bahuvīhi* compound). But a compound may also have two heads. These compounds are called coordinate compounds or *dvandva* compounds. For example, the word meaning 'parents' in many languages is 'mother-father' or 'father-mother' (compare Chinese 父母 *fumu* 'father-mother'). In this kind of dvandva compound, the referent of the whole is the UNION (or event superset of the union) of the referents of the parts. There is another kind of dvandva compound, common in English, in which the referent of the whole is the INTERSECTION of the referents of the parts. Consider *singer-songwriter*, *fighter-bomber*, and *scientist-supervillain-entrepeneur*.

Compounds that have only one heads can be divided into two types: SUBORDINATE COMPOUNDS and ATTRIBUTIVE COMPOUNDS. In subordinate compounds, the modifier (the non-head) is an ARGUMENT of the head. For example, in *dog catcher*, *dog* is an argument of *catcher* (a person who **catches** dogs). In attributive compounds, the modifier specifies some attribute(s) that apply to the head. For example, a *dogfight* does not involve fighting a dog, but rather fighting like a dog (in aerial combat).

In both of these types of compounds, none of the words modifies any of the others (or serves as an exclusive head). It is as if all of them are heads

equally.

Compounding and affixation

Consider the following examples from Hmong[1] All of these words were originally compounds, and the word at the left was the HEAD of the compound, meaning that it was the part of the compound that contributed the basic meaning of the compound. Originally, *kev-phem* meant 'bad way' and *tub-ngag* meant 'steal(ing) boy'. However, as the number of constructions like this grew, the meanings of *kev* and *tub* became more general (until *kev* was similar to the English suffix *-ness* and *tub* was similar to the English suffix *-er* as in *help-er*. At some point, *kev* and *tub* have stopped being independent words (when they occur in this construction) and have become derivational prefixes. However, it is impossible to pinpoint the exact point in time at which this threshold was crossed.

FORM	MEANING (AS COMPOUND)	TRANSLATION
kev-noj	way-eat	'eating'
kev-haus	way-drink	'drinking'
kev-kaaj	way-bright	'brightness'
kev-zoo	way-good	'goodness'
kev-phem	way-bad	'evil'
kev-kawm	way-study	'studying'
tub-txib	boy-send	'messenger'
tub-khaiv	boy-send	'servant'
tub-nyag	boy-steal	'thief'
tub-ncig	boy-be_around	'funeral helpers'

[1] Hmong is a minority language of China, Vietnam, Laos, and Thailand. As far as linguistics can determine, it is not related to any major languages.

Table 2: The derivational prefixes *kev-* and *tub-* developed from compounds where the heads were 'way' and 'boy', respectively.

Similarly, there is a large body of word-forming morphology in English that could either be understood as compounding or derivation. So-called neoclassical compounds, which are widely used in scientific and technical terminology, consist of morphemes borrowed from Greek and Latin. Take for example *hypnopaedia* 'learning through sleep' which consists of the two morphemes *hypno-* 'sleep' and *-paedia* 'learning'. *hypno-* looks, in some ways, like a prefix—it cannot really occur by itself (at least in formal speech) and is always preposed to a base. However, it has a concrete meaning rather than an abstract one, and affixes typically have abstract meanings.

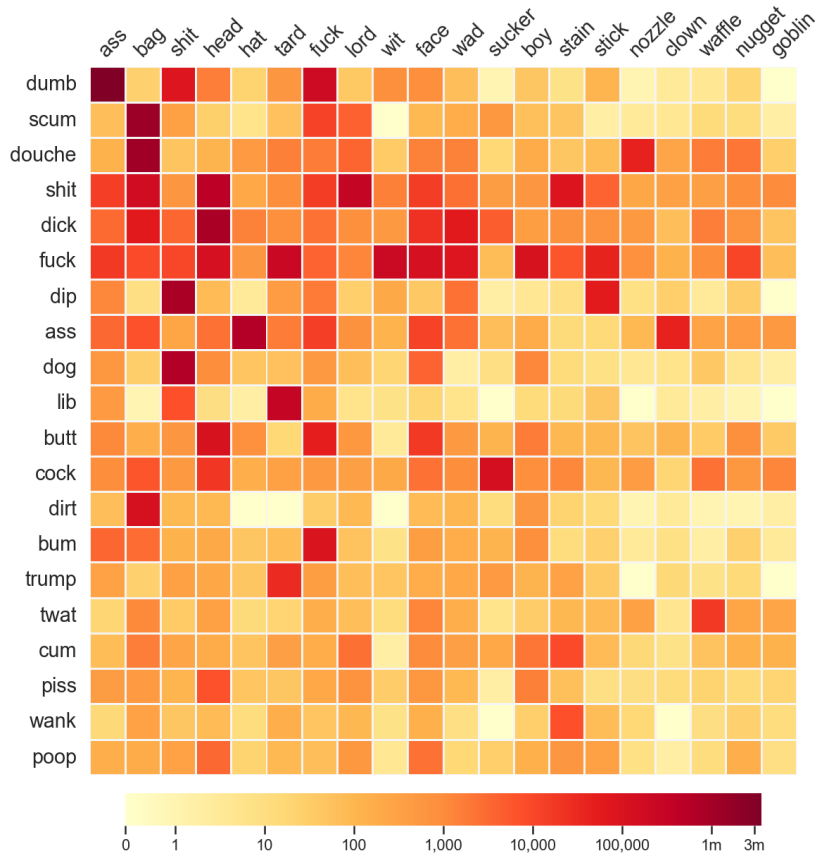

Figure 1: Pejorative compounds from Reddit posts (`https://colinmorris.github.io/blog/compound-curse-words`).

Compounds and Productivity

While not all compounding constructions are highly productive, compounds tend to be among the most productive of word-formation processes. Consider Figure 1, depicting the occurrence of various words in pejorative compounds scraped from Reddit. The color in each cell indicates the absolute frequency of the corresponding subordinate compound (modifiers in rows, heads in columns). Note that some of the compounds are much more common than others (e.g., *dumbass*) but that almost all of the possible combinations occur. Exceptions include *dipgoblin* and *scumwit*. These are easily ruled out on semantic grounds, suggesting that—from a grammatical point of view—this kind of compounding is exceptionally productive in English.

Derivation

Derivation is also a common word-formation strategy. As mentioned above, it's hallmarks are changing meaning (the concept), changing part-of-speech, and governing scopes. The last of these deserves some explanation. In math-

ematics, $(1 + 2) \times 3$ has a different value than $1 + (2 \times 3)$ even though the numbers and operations because in the first case 3 has scope over $1 + 2$ and in the second case, 1 has scope over 2×3.

Now consider the following examples from the Bantu language Bemba:

(4) a. Na- a- mon -an -y -a Mwape na Mutumba
1SG- PAST- see -REC -CAUS -FV 1.Mwape and 1.Mutumba
'I made Mwape and Mutumba see each other.'

b. Mwape na Chilufya ba- a- mon -eshy -an -a -FV
1.Mwape and 1.Chilufya 2SG.MASC- PAST- see -CAUS -REC
'Mwape and Chilufya made each other see Mutumba.'

1SG	first-person singular subject
2SG.MASC	second person singular masculine subject
PAST	past tense
REC	reciprocal (X each other)
CAUS	causitive (cause X to happen)
FV	final vowel (suffix with complex grammar)

Table 3: Explanation of glosses from Bemba examples.

To unpack things a bit, see Table 3.

In V-REC-CAUS, where CAUS has scope over REC, the meaning is 'cause to V each other' but V-CAUS-REC means 'cause each other to V.' This kind of effect, in which you can obtain different meanings by changing the sequence of two morphemes, is very common in derivation but is not common in inflection (in which the ordering of morphemes is typically fixed following general principles). We will discuss more about scope and affix ordering in the next lecture.

Deriving Nouns

The most common and diverse set of derivational operations convert words into nouns (especially verbs, adjectives. and other nouns). Some illustrative examples are presented in Table 4.

DEVERBAL NOUN	English	*drink*		*drinker*	
	Arabic	*ħamala*	'carry'	*ħammaal*	'carrier'
	English	*invite*		*invitee*	
	Spanish	*picar*	'mince'	*picadora*	'meat grinder'
	Russian	*otkryt*	'discover'	*otkrytie*	'discovery'
DEADJECTIVAL NOUNS	Japanese	*atarasii*	'new'	*aratasisa*	'newness'
	Russian	*umnyj*	'clever'	*umnik*	'clever guy'
DENOMINAL NOUNS	Spanish	*gato*	'cat'	*gatito*	'little cat'
	Russian	*boroda*	'beard'	*borodišča*	'huge beard'
	English	*child*		*childhood*	
	Arabic	*Miṣr*	'Egypt'	*miṣriyyu*	'Egyptian'
	German	*König*	'king'	*Königin*	'queen'

Table 4: Derived nouns (adapted from Haspelmath and Sims (2010).

Deriving Verbs

Derivational operations that covert words into verbs are also quite common (but are less common and less diverse than those that derive nouns):

DEVERBAL VERBS	Korean	*cwuk*	'die'	*cwuki*	'kill'
	German	*laden*	'load'	*beladen*	'load onto'
	Swedish	*öppna*	'open (tr.)'	*öppnas*	'open (intr.)
	Greenlandic	*sini-*	'sleep'	*sinikkuma-*	'want to sleep'
	English	*write*		*rewrite*	
	Swahili	*choma*	'stick in'	*chomoa*	'pull out'
DENOMINAL VERBS	Spanish	*pirata*	'pirate'	*piratear*	'pirate (verb)'
	English	$bottle_N$		$bottle_V$	
	Russian	*sol*	'salt'	*solit*	'salt (verb)'
DEADJECTIVAL VERBS	Russian	*černyj*	'black'	*černit*	'make black'
	Spanish	*verde*	'green'	*verdear*	'become green'

Table 5: Derived verbs (adapted from Haspelmath and Sims (2010).

One special class of morphological operations is exemplified by Swedish *öppnas* 'open' (as in 'the door opened'). This kind of morphology, called VALENCY CHANGING morphology, changes the number of ARGUMENTS that a verb can take. In the Swedish case, it derives an intransitive verb (which takes only a subject) from a transitive verb (which takes a subject and an object). We saw some other instances in the Bemba example at the beginning of this session: CAUSATIVES and RECIPROCALS. However, there are a number of other such categories. The best know valency reducing operation is the PASSIVE (familiar because busybodies criticize you for using it in English prose). Other examples, covered in Chapter 11 of Haspelmath and Sims include ANTICAUSATIVE, ANTIPASSIVE, and APPLICATIVES.

Deriving Adjectives and Adverbs

Many languages also have morphology for deriving verbs and adjectives. Examples are provided in Table 6.

Inflection

Case

The most prototypical kind of inflection is probably case. Nouns, DETERMINERS (like articles), and adjectives are inflected for case based on the grammatical role (within the sentence) of the phrase that they are in. English only has case marking in pronouns:

(5) a. She saw him.

DEVERBAL ADJECTIVES	Basque	*jan*	'eat'	*jangarri*	'edible'
	Spanish	*hablar*	'talk'	*hablador*	'talkative'
DENOMINAL ADJECTIVES	Russian	*korolʲ*	'king'	*korolevskij*	'royal'
	Ponapean	*pihl*	'water'	*pilen*	'watery'
	Russian	*voda*	'water'	*bezvodnyj*	'waterless'
	German	*Kupfer*	'copper'	*kupfern*	'made of copper'
DEADJECTIVE ADJECTIVES	Tzutujil	*kaq*	'red'	*kaqkoj*	'reddish'
	Turkish	*yeni*	'new'	*yepyeni*	'brand new'
	German	*schön*	'beautiful'	*unschön*	'ugly'

Table 6: Derived adjectives (adapted from Haspelmath and Sims (2010).

b. * Her saw him.
c. * She saw he.

The NOMINATIVE case (like *she* and *he*) appears in noun phrases that are the SUBJECT of a CLAUSE. The ACCUSATIVE case occurs in noun phrases that are the OBJECT of a clause. GENITIVE case occurs in noun phrases that are possessing another noun phrase and DATIVE case occurs in noun phrases that are recipients, goals, etc., of a clause. Some languages do not have morphological case at all and some have many more than these.

Person, Number, and Gender

In many languages, verbs change their morphological form based on the PERSON, NUMBER, and GENDER of their subjects (and sometimes objects). Adjectives, likewise, change their form based on the number and gender of the nouns they modify. Person refers to who/what the subject (or object) of the sentence is relative to the participants in the conversation. The most common system is as in English pronouns (first, second, and third persons; *I/me/we*, *you*, and *he/she/it/they/them*). However, English has a very impoverished agreement system:

(6) a. She run-s.
b. They ran.
c. I run.
d. * She ran-s.
e. * They ran-s.
f. * I run-s.

The other properties that are involved are TENSE (for which, see below) and number. The most common number system is singular versus plural, though many languages also have a dual number. Some even have *paucal* for small numbers greater than two.

English used to have dual number in addition to singular and plural

Unlike English, many languages have grammatical gender. They divide nouns into somewhat arbitrary classes. In European languages, these are usually called MASCULINE, FEMININE, and NEUTER (because words

with a particular cultural gender as assigned to a particular class). However, many languages have different systems of noun classes that play a role in agreement. In these systems, adjectives typically change form depending on the class/gender of the noun they modify and verbs change form depending on the class/gender of their subject (and sometimes object).

Tense, Aspect, and Modality

TENSE marks a verb based on when the event signified by the verb phrase occurred relative to the present. Languages divide up time in different ways. English, for example, has a simple PAST/NON-PAST system (morphologically). We contrast *walk* (non-past) and *walked* (past) but have to add something special to make a future *will walk*. Aspect, instead, is about the structure of an event in time. For example, PERFECTIVE aspect presents an event as a whole whereas IMPERFECTIVE aspect presents it as ongoing. The English PROGRESSIVE (as in *walking*) is a kind of imperfective. Modality concerns the attitude of the language user relative to the event being signified—whether it is actual or hypothetical, for example, or imperative. The SUBJUNCTIVE is a kind of MOOD (and mood is a subset of modality).

Tense, and especially aspect, and most particularly mood are unlike case, person, number, and gender in that they actually seem to change the meaning of sentences. Why call them inflectional then? The answer is that they have other properties of inflection. They participate, in a limited way, in the syntactic structure of languages, they are almost never scopal, and they are paradigmatic (see below).

Voice

VOICE is the least inflection-like of the inflectional properties discussed here. It concerns how syntactic rules are aligned with semantic functions and is marked on verbs. Consider the following examples from English:

(7) a. Tamar recoils in horror when her papa **kisses** her on the cheek.
b. Tamar recoils in horror when she is **kissed** on the cheek by her papa.

In 7a, *her papa* is the subject of the sentence and *her* is the object. The verb form is *kisses*. But in 7b, the verb form is *kissed*, the subject is *she*, and *her papa* has been demoted to what is called an OBLIQUE.

The passive voice very clearly changes the meaning of verbs. In some languages, there is no reason to treat voice (valency changing morphology) as inflection. However, in other languages (like Japanese, Greek, and Latin), voice has all of the other characteristics of inflection. Furthermore, it clearly has relevance to the syntax.

Paradigms, Cumulative Exponence, and Extended Exponence

One of the key characteristics of inflectional morphology is its paradigmatic nature. A paradigm is a kind of matrix with one or more dimensions. Each dimension corresponds to a MORPHOSYNTACITC PROPERTY (like case, number, gender, TAM, or voice) and each row or column in the matrix corresponds to a value of that feature, or a VALUE.

The English noun paradigm is simple—it has just one feature with just two values:

Table 7: The paradigm for English SEA.

SINGULAR	sea
PLURAL	seas

However, many languages have very complicated noun paradigms. A somewhat more complicated example can be found in Latin, with two features, case and number: The case feature can take any of six values, and

Table 8: The paradigm for Latin DOMINUS 'lord; master'.

	SINGULAR	PLURAL
NOMINATIVE	dominus	domini
VOCATIVE	domine	domini
ACCUSATIVE	dominum	dominos
GENITIVE	domini	dominorum
DATIVE	domino	dominis
ABLATIVE	domino	dominis

there is a word form for every one of them. As in English, the number feature/property takes only two values. A combination of such values (e.g., DATIVE PLURAL) is called a MORPHOSYNTACTIC CATEGORY.

Table 9: The paradigm for Latin DOMINUS 'lord; master' with the suffixes segmented.

	SINGULAR	PLURAL
NOMINATIVE	domin-us	domin-i
VOCATIVE	domin-e	domin-i
ACCUSATIVE	domin-um	domin-os
GENITIVE	domin-i	domin-orum
DATIVE	domin-o	domin-is
ABLATIVE	domin-o	domin-is

(8) Two observations about the Latin paradigm:

a. A single suffix often realizes multiple properties. *-orum* realizes GENITIVE and PLURAL. This is called CUMULATIVE EXPONENCE.

b. The same suffix may realize multiple categories. *-o* realizes both DATIVE SINGULAR and ABLATIVE SINGULAR. This is called SYNCRETISM.

These are characteristics that follow from the paradigmatic nature of inflection and which are generally absent in derivation.

There is another such characteristic, and that is EXTENDED EXPONENCE. Consider the Classical Greek verb ελελῠ́κᾰτε (*elelýkete*), the second-person plural active past perfect indicative form of λῡ́ω (*luo*) 'unfasten':[3] The

[3] P. H. Matthews. *Morphology*. Cambridge University Press, Cambridge, 1974

The *ly* in this word is the same *ly* in *analysis*.

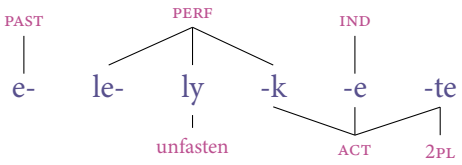

Figure 2: Extended exponence in Greek *elelykete*. PAST = past tense, PERF = perfect aspect, IND = indicative modality, ACT = active voice, and 2PL = second personal plural subject agreement.

parts of this word that signify perfect and indicative are "spread out" across multiple MORPHS (parts of the word). No one morph signifies PERF, but three different morphs are what they are because the verb has perfect aspect: the root is reduplicated, the root's vowel is shortened, and *-k* is suffixed to the root. [4]

[4] But *-k* only occurs in perfect verbs when they are also active.

Not all languages are like Greek and Latin. Some languages are largely without cumulative and extended exponence. These languages are said to have a LOW DEGREE OF FUSION whereas languages like Greek and Latin are said to have a HIGH DEGREE OF FUSION. However, when languages have fusion of this kind, it is overwhelmingly in inflection and is a product of its paradigmatic nature.

Implications

Hypothesis: The difference between compounding, derivation, and inflection is really a matter of information context. I posit that roots are simply very informative morphemes (high surprisal). Derivational affixes are somewhat less informative. Inflectional affixes are minimally informative (low surprisal). Given a corpus segmented into morphemes, this could be investigated empirically. **This would make a good paper topic.**

Also, there are implications for tokenization:

(9) **Revised Optimal Slicing Hypothesis**

a. An optimal tokenization segments inflectional morphemes from bases but not derivational morphemes

b. An optimal tokenization segments inflection by morphological category, not by morph, in languages with high degrees of fusion (corollary: no more than one slice is best)

c. However, optimal tokenization segments compounded words/stems

The motivation for this hypothesis, and also some refinements, will be presented in the next lecture.

References

Martin Haspelmath and Andrea Sims. *Understanding Morphology*. Hodder Education, London, 2nd edition, 2010.

P. H. Matthews. *Morphology*. Cambridge University Press, Cambridge, 1974.