

Contents

	Acknowledgments	xvii
1	Introduction/motivation	1
	#0 Knowing about linguistic structure is important for feature design and error analysis in NLP.	1
	#1 Morphosyntax is the difference between a sentence and a bag of words.	2
	#2 The morphosyntax of a language is the constraints that it places on how words can be combined both in form and in the resulting meaning.	3
	#3 Languages use morphology and syntax to indicate who did what to whom, and make use of a range of strategies to do so.	5
	#4 Languages can be classified ‘genetically’, areally, or typologically.	5
	#5 There are approximately 7,000 known living languages distributed across 128 language families.	7
	#6 Incorporating information about linguistic structure and variation can make for more cross-linguistically portable NLP systems.	8
2	Morphology: Introduction	11
	#7 Morphemes are the smallest meaningful units of language, usually consisting of a sequence of phones paired with concrete meaning.	11
	#8 The phones making up a morpheme don’t have to be contiguous.	11
	#9 The form of a morpheme doesn’t have to consist of phones.	13
	#10 The form of a morpheme can be null.	13
	#11 Root morphemes convey core lexical meaning.	14
	#12 Derivational affixes can change lexical meaning.	16
	#13 Root+derivational affix combinations can have idiosyncratic meanings.	17
	#14 Inflectional affixes add syntactically or semantically relevant features.	18
	#15 Morphemes can be ambiguous and/or underspecified in their meaning.	19
	#16 The notion ‘word’ can be contentious in many languages.	20
	#17 Constraints on order operate differently between words than they do between morphemes.	21
	#18 The distinction between words and morphemes is blurred by processes of language change.	22

#19 A clitic is a linguistic element which is syntactically independent but phonologically dependent.	23
#20 Languages vary in how many morphemes they have per word (on average and maximally).....	24
#21 Languages vary in whether they are primarily prefixing or suffixing in their morphology.	25
#22 Languages vary in how easy it is to find the boundaries between morphemes within a word.	26
3 Morphophonology	29
#23 The morphophonology of a language describes the way in which surface forms are related to underlying, abstract sequences of morphemes.	29
#24 The form of a morpheme (root or affix) can be sensitive to its phonological context.	29
#25 The form of a morpheme (root or affix) can be sensitive to its morphological context.	31
#26 Suppletive forms replace a stem+affix combination with a wholly different word.	32
#27 Alphabetic and syllabic writing systems tend to reflect some but not all phonological processes.....	33
4 Morphosyntax	35
#28 The morphosyntax of a language describes how the morphemes in a word affect its combinatoric potential.	35
#29 Morphological features associated with verbs and adjectives (and sometimes nouns) can include information about tense, aspect and mood.	36
#30 Morphological features associated with nouns can contribute information about person, number and gender.	38
#31 Morphological features associated with nouns can contribute information about case.	40
#32 Negation can be marked morphologically.	41
#33 Evidentiality can be marked morphologically.	42
#34 Definiteness can be marked morphologically.	43
#35 Honorifics can be marked morphologically.	43
#36 Possessives can be marked morphologically.	44
#37 Yet more grammatical notions can be marked morphologically.....	46

#38 When an inflectional category is marked on multiple elements of sentence or phrase, it is usually considered to belong to one element and to express agreement on the others.	46
#39 Verbs commonly agree in person/number/gender with one or more arguments.	47
#40 Determiners and adjectives commonly agree with nouns in number, gender and case.	48
#41 Agreement can be with a feature that is not overtly marked on the controller.	49
#42 Languages vary in which kinds of information they mark morphologically. ..	50
#43 Languages vary in how many distinctions they draw within each morphologically marked category.	51

5 Syntax: Introduction 53

#44 Syntax places constraints on possible sentences.	53
#45 Syntax provides scaffolding for semantic composition.	54
#46 Constraints ruling out some strings as ungrammatical usually also constrain the range of possible semantic interpretations of other strings.	54

6 Parts of speech 57

#47 Parts of speech can be defined distributionally (in terms of morphology and syntax).	57
#48 Parts of speech can also be defined functionally (but not metaphysically). ...	58
#49 There is no one universal set of parts of speech, even among the major categories.	59
#50 Part of speech extends to phrasal constituents.	60

7 Heads, arguments and adjuncts 61

#51 Words within sentences form intermediate groupings called constituents. ...	61
#52 A syntactic head determines the internal structure and external distribution of the constituent it projects.	63
#53 Syntactic dependents can be classified as arguments and adjuncts.	65
#54 The number of semantic arguments provided for by a head is a fundamental lexical property.	65
#55 In many (perhaps all) languages, (some) arguments can be left unexpressed. .	66
#56 Words from different parts of speech can serve as heads selecting arguments .	67
#57 Adjuncts are not required by heads and generally can iterate.	69

#58 Adjuncts are syntactically dependents but semantically introduce predicates with take the syntactic head as an argument.	69
#59 Obligatoriness can be used as a test to distinguish arguments from adjuncts. .	71
#60 Entailment can be used as a test to distinguish arguments from adjuncts....	71
#61 Adjuncts can be single words, phrases, or clauses.	72
#62 Adjuncts can modify nominal constituents.	73
#63 Adjuncts can modify verbal constituents.	73
#64 Adjuncts can modify other types of constituents.	74
#65 Adjuncts express a wide range of meanings.	74
#66 The potential to be a modifier is inherent to the syntax of a constituent.	74
#67 Just about anything can be an argument, for some head.	75
8 Argument types and grammatical functions	79
#68 There is no agreed upon universal set of semantic roles, even for one language; nonetheless, arguments can be roughly categorized semantically.	79
#69 Arguments can also be categorized syntactically, though again there may not be universal syntactic argument types.	80
#70 A subject is the distinguished argument of a predicate and may be the only one to display certain grammatical properties.	83
#71 Arguments can generally be arranged in order of obliqueness.	84
#72 Clauses, finite or non-finite, open or closed, can also be arguments.	85
#73 Syntactic and semantic arguments aren't the same, though they often stand in regular relations to each other.	86
#74 For many applications, it is not the surface (syntactic) relations, but the deep (semantic) dependencies that matter.	88
#75 Lexical items map semantic roles to grammatical functions.	88
#76 Syntactic phenomena are sensitive to grammatical functions.	90
#77 Identifying the grammatical function of a constituent can help us understand its semantic role with respect to the head.	91
#78 Some languages identify grammatical functions primarily through word order.	91
#79 Some languages identify grammatical functions through agreement.	93
#80 Some languages identify grammatical functions through case marking.	95
#81 Marking of dependencies on heads is more common cross-linguistically than marking on dependents.	97
#82 Some morphosyntactic phenomena rearrange the lexical mapping.	97

9	Mismatches between syntactic position and semantic roles	101
	#83 There are a variety of syntactic phenomena which obscure the relationship between syntactic and semantic arguments	101
	#84 Passive is a grammatical process which demotes the subject to oblique status, making room for the next most prominent argument to appear as the subject.	101
	#85 Related constructions include anti-passives, impersonal passives, and middles.	103
	#86 English dative shift also affects the mapping between syntactic and semantic arguments.	104
	#87 Morphological causatives add an argument and change the expression of at least one other.	106
	#88 Many (all?) languages have semantically empty words which serve as syntactic glue.	107
	#89 Expletives are constituents that can fill syntactic argument positions that don't have any associated semantic role.	109
	#90 Raising verbs provide a syntactic argument position with no (local) semantic role, and relate it to a syntactic argument position of another predicate.	110
	#91 Control verbs provide a syntactic and semantic argument which is related to a syntactic argument position of another predicate.	112
	#92 In complex predicate constructions the arguments of a clause are licensed by multiple predicates working together.	113
	#93 Coordinated structures can lead to one-to-many and many-to-one dependency relations.	115
	#94 Long-distance dependencies separate arguments/adjuncts from their associated heads	116
	#95 Some languages allow adnominal adjuncts to be separated from their head nouns.	118
	#96 Many (all?) languages can drop arguments, but permissible argument drop varies by word class and by language.	119
	#97 The referent of a dropped argument can be definite or indefinite, depending on the lexical item or construction licensing the argument drop.	121
10	Resources	123
	#98 Morphological analyzers map surface strings (words in standard orthography) to regularized strings of morphemes or morphological features.	123
	#99 'Deep' syntactic parsers map surface strings (sentences) to semantic structures, including semantic dependencies.	124

#100 Typological databases summarize properties of languages at a high level. . .	125
Summary	125
A Grams used in IGT	127
Bibliography	131
Author's Biography	153
General Index	155
Index of Languages	165