

***The Sixth
Annual***

**North American
Computational
Linguistics
Olympiad**

www.naclo.cs.cmu.edu

**INVITATIONAL ROUND
(SOLUTIONS)
March 13, 2012**

I. The silence of the goats (I/I)

When Esperanto speakers want to specify the time of an event simply with respect to current time, they use simple tenses:

Past: mangis (ate), Present: mangas (eat), Future: mangos (will eat) When they want to specify the time of an event with respect to another (anchor) point in time they use a suffix and an ending. The ending, taken from the simple form indicates the anchor time. The suffix indicates past, present or future with respect to the anchor time. The form of the verb is Root-X(n)t-Ys, where: X and Y are a, o, or l for present, past, and future respectively.

X(n)t is a suffix that indicates the time of the event in relation to the anchor time expressed by the ending Ys. Ys is the ending which indicates the anchor point in time (in relation to the present moment) The absence of n in a X(n)t suffix indicates passive voice; its presence indicates active voice.

The time relations shown by the morphemes used in the material:

o-(n)t — future	o-s - future
i-(n)t — past	i-s - past
a-(n)t — ?	a-s - present

8. La kapro manĝontos. ont+os = future in relation to future, active voice.

The goat will be going to eat = The goat will be about to eat .

9. La kapro manĝitas. it+as = past in relation to present, passive voice

The goat has been eaten = The goat was eaten.

10. The goat was eating. Present in relation to past, active voice. In the material, there is no suffix indicating the present tense in relation to another point in time, but it can be detected using the table above – a(n)t.

La kapro manĝantis.

11. The goat is being eaten. Present in relation to present, passive voice.

La kapro manĝatas.

Comment: the fact is, that the complex forms used in the problem, which are only used in spoken Esperanto, are shortened from the standard form **esti** (to be) + **participle**, f.e. „**La kapro manĝontis**” is shortened from **La kapro estis manĝonta** (was + future participle)

(with contributions from Anatole Gershman)



J. Ik heb voorspeld (I/I)

J-1

1 gedeeld	2 gehoord	3 getapt	4 verhuisd	5 geland
6 geklopt	7 gemokt	8 gerookt	9 gerot	10 getobd

J-2. The reason that deriving the plain form from the past participle is not possible in all cases for this set of data is the fact that verb stems ending in a double/compound consonant (e.g., *kunn-en*) as well as stems ending in a consonant cluster consisting of a C and d (e.g., *brand-en*) create past participles with a final consonant cluster of the form Cd. Therefore, deriving backwards, a past participle form ending in -nd (e.g., *gebrand*) could derive either from a plain form with an -nd cluster (e.g., *branden*) or a plain form with a double/compound consonant (e.g., **brannen*). Similarly, *geland* may derive from *landen* or **lannen* and *gekund* may derive from *kunden* or **kunnen*.

All other hypothetically plausible ambiguities are not supported by examples in the given data and were, therefore, not admissible. For example, another ambiguity concerns a plain form with a double vowel. Evidence from long vowel plain forms is that the double vowel would stay the same, and this is indeed the case (e.g., *leeren* 'to learn' becomes *geleerd*). Therefore, a double vowel in the past participle could have come from a double vowel in the plain form. However, as there are no examples of this in the data, this is not an admissible example.

(With contributions from Aleka Blackwell)



K. A fox among the h (I/I)

K-1 BE, KIWUS, FLI, FLEUM, GEES

K-2 Dumutche's program doesn't stop applying rules when one of them succeeds – it keeps looking for applicable rules and then applies them to the output of the previous rules.

This gives the right output for "walruses" and "foxes" – it removes the "s", then continues on and removes the "e" – but goes very wrong with "horses", "hens", etc.

K-3 We can determine that:

"Remove S" must come before "Remove E"; otherwise, we would get WALRUSE, FOXE, MOU, etc. instead.

"Remove S" must come before "Remove EN"; if it came after, we would get HEN instead.

"Remove S" must come before "Replace IES with Y"; if it came after, we would get GUPPY instead.

"Remove S" must come before "Replace I with US"; if it came after, we would get FUNGU instead.

"Replace I with US" must come before "Remove E"; if it came after, we would get GUPPUS instead.

"Remove E" must come before "Replace A with UM"; if it came after, we would get ALGA instead.

"Replace ICE with OUSE" must come before "Remove E"; if it came after, we would get MIC instead.

Rules do not apply twice – that is, Dumutche's program probably goes through each rule in the list exactly once and only goes through the list once. Otherwise, we would get things like CHIMPANZ or WALRU, etc.

K-4 There is no one order of rules that will make Dumutche's program work, for several reasons:

1. The feeding of "Remove E" by "Remove S" is necessary to get "walrus" and "fox" correct, but it's exactly this interaction that produces "hors" and "chimpanzee."

2. No order will correctly produce "mouse." Consider the two rules (A) "Remove E" and (B) "Replace ICE with OUSE". If A comes before B, we get "mic"; if B comes before A, we get "mous" (or even "mou").



L. Who is good? (I/I)

L-1 (accept either 'the' or 'a').

'The man walks home.'

'Does the woman hit my father?' or 'Does my father see the woman?'

'Who is not a liar?'

'Does the teacher see the woman?'

L-2 (accept any word order).

[hu:ʔunikat ʂu tʃipomkat]

[ʂuŋa:li to:wq hu:ʔunikat]

[ʔivi nawitmal qaj nona:ji to:wq]

[haxʂu polo:v) or [hax ʂu polo:v]



M.The Deschamps Codice (I/I)

M-1

1-3: 1st, 2nd and 3rd consonant of the surname. If there are not enough consonants, add the first vowel.

4-6: 1st, 3rd and 4th consonant of the given name. If there are only 3 consonants, take the 1st, the 2nd and the 3rd one. If there are not enough consonants, add the first vowel.

7-8: Year of birth (last two digits)

9: Month of birth (A = January, C = March, E = May, H = June, L = July, P = September, R = October, T = December). Even though it is unclear why some letters are not used, it is easy to notice that the letters are ordered alphabetically, which allows to establish the missing correspondences (B = February, D = April, M = August, S = November)

10-11: Day of birth (for a female, add 40)

12: Z

13: Continent (1 = Europe, 2 = Asia, 3 = Africa, 4 = North America, 6 = South America)

14-15: Country within a continent. Countries are ordered alphabetically, assuming that their Italian names are similar to the English ones (Europe: 10 = France / *Francia*, 15 = Greece / *Grecia*, 33 = Switzerland / *Svizzera*; South America: 00 = Argentina / *Argentina*, 01 = Bolivia / *Bolivia*, 03 = Chile / *Cile*). The most likely candidate for 02 in South America is Brazil / *Brasile*.

16: Checksum (algorithms for counting it are too difficult to be deduced from the problem)

M-2

Y is regarded as a consonant, no matter how it is actually pronounced.

First and Last Name	Sex	Date of birth	Country of birth	Codice fiscale
Pedro Santos	M	Feb 03, 1986	Brazil	SNTPDR86B03Z602Z
Rita Uwak	F	Apr 28, 1960	A country in Africa	WKURTI60D68Z335A
Yukari Miyama	F	Aug 30, 1990	A country in Asia	MYMYKR90M70Z219U

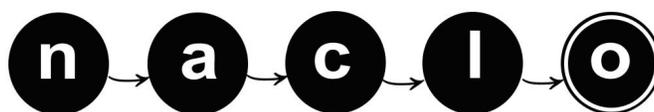
M-3

The solvers may provide other given names and surnames. The grader has to check whether the first six letters of the *codice fiscale* can be generated from them correctly (disregarding the checksum). Z335 is actually Nigeria, and Z219 is Japan.



N. Waanyi (1/2)

1.	I am sitting in the camp.
2.	Then they both ask(ed) this man.
3.	Now I am not hungry.
4.	Jungku bula nawunu burrurri, kirriya.
5.	Jarrba nyulu nanangkani kirriyaa kaku.
6.	Dabarraba nyulu nana waliji nangkani burrurrii karrinjana kundana.



N. Waanyi (2/2)

This is an exercise in exploring the structure of a language with very different grammatical rules from English. The real challenges lie in the syntax, as explained below.

- Finding the lexical correspondences (i.e. the dictionary words or vocabulary) is relatively easy, especially when you recognise that some words contain a long suffix (e.g. *kirriyawurru* in 7 contains *kirriya* in 1). Fortunately the morphology is very straightforward, so endings are easy to recognise; but pay attention to small differences such as that between *kirriya* in 1 and *kirriyaa* in 9.
- The first syntactic problem concerns words like *nyulu* and *bula*, which don't seem to correspond systematically to anything in the English. Almost every sentence contains one of these words (or some other bisyllabic non-vocabulary word: *yalu*, *nayi*, *ninji*, *ngawu*), so you might suspect (rightly) that they're like auxiliary verbs; so let's call them auxiliaries. But what distinction do these verbs indicate? You might consider alternatives such as tense or positive/negative, but it turns out in fact that it's the 'person' and number of the sentence's 'actor' (more on this below): *ngawu* = 'I', *ninji* = 'you', *bula* = 'the two of them', *yalu* = 'they', *nyula* = 'he/she/it'. (*Nayi* in 24 is a puzzle; maybe it just means 'there isn't'.)
- The second syntactic problem is word order. One regularity is that the auxiliary is very often the second word in the sentence, especially if you ignore on the one hand introductory material separated from the rest of the sentence by a comma (e.g. 6), and on the other the word *barri*, which means 'then'. With these two exclusions, there are no exceptions: the auxiliary always follows just one word. Moreover, that one word is very often the main verb of the sentence – but not always. For instance, 31 starts *rajiwurru barri bula*, 'they both returned then', but 32 moves the verb *rajiwurru* after the auxiliary: *Balikajini bula kannga rajiwurru*, 'hungry they both return'. Apart from these two rules, the order of elements in a sentence is free.
- The third problem is that nouns have more than one form which varies from sentence to sentence; for instance, 'woman' is *kirriya* in 1 and 5, but *kirriyaa* in 9 and 10. In some examples the extra material seems to express the same meaning as an English preposition; for example, *kirriyawurru* is translated as 'to the woman' (combined with *kanungku*, translated as 'approaching' in 19 and 21). But this doesn't help with the variation between *kirriya* and *kirriyaa*, or the similar alternation between *burrurri* and *burrurrii* for 'man' in 2 and 4. You may think it's like the 'case' changes found in languages like German (or in English pronouns, such as *he* and *him*), but that contrasts grammatical subjects and objects (e.g. *he came*, and *she saw him*).
- This pattern is called a 'nominative-accusative' case system, where 'nominative' and 'accusative' are the traditional names for the cases used as subjects and objects. That's not what's happening in Waanyi, where the short form is used both as subject in 1 or 2 (X is standing/sitting ...) and also as object in 17 and 19 (... saw X). Waanyi follows a completely different system, called a 'nominative-ergative' case system, where a special form (in Waanyi, the longer one) is used for the 'actor' in a two-part action, where the actor does something to someone or something else – in other words, for the subject of a verb that also has an object. For instance, 'woman' is normally *kirriya*, but in 9 it's *kirriyaa* because it means 'that woman takes that meat ...' and likewise in 10 meaning 'that woman eats that meat'.



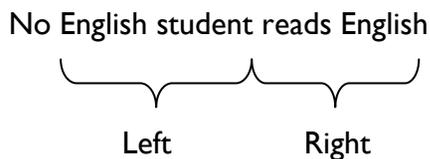
O. The Little Engine that Could... Read (I/I)

O-1. C. (The illegitimate inference is “No English student can read Russian” => “No student can read Russian.” We certainly hope that Russian students, for example, can read Russian!)

O-2. Every teacher can read.

- O-3.** G. Upward
H. Upward
I. Downward
J. Downward
K. Upward
L. Downward

The terms “left” and “right” determine which of the two “arguments” of the quantifier (the first one or the second one, respectively) the rule applies to.



The terms “upward” or “downward” indicate whether a *more general* or *more specific* replacement is warranted. (So “English teacher” is more specific than “teacher”, “reads English” is more specific than “reads”, “reads English and Russian” is more specific than “reads English”, etc.)

So a “left upward” quantifier allows the replacement of the “left” argument of the quantifier with a more general phrase, and a “right downward” quantifier allows the replacement of the “right” argument with a more specific phrase.



P. Jeg kan tælle (1/2)

P-1	A	seks	6
	B	nioghalvtreds	59
	C	treogtyve	23
	D	femoghalvfems	95
	E	toogtres	62
	F	halvfjerds	70

P-2	A	syv	7
	B	fireoghalvtreds	54
	C	enogtyve	21
	D	femogfirs	85
	E	nioghalvfems	99

Danish numbers greater than 10 express the units first, then the tens, linked by *og* 'and'. Although in the original data only one element (*ni*, 9) is repeated, the pattern can be seen further in the numbers you are asked to translate. For example, you know *tre* is 3 and *toogtyve* is 22, so *treogtyve* shouldn't be too hard to guess.

For most of the answers it is just a matter of decomposing the given examples and reconstructing them.

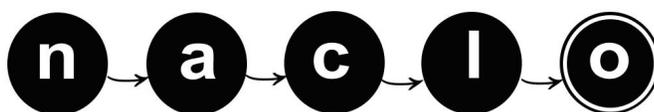
These numerals can be worked out from data:

	1	2	3	4	5	6	7	8	9
units	en	to	tre	fire	fem	seks	syv		ni
tens		tyve			halvtreds	tres	halvfjerds	firs	

There are two elements not given in the data: 90, and *halvfems*. If we ask for a word not included in the data, it must be the case that you are expected to be able to work it out.

You have to notice a pattern for the tens other than 20: the 'even' tens 60 and 80 are expressed as $n \times 20$ (cf English *score*): *tre+s* 'three score', *fir(e)+s* 'four score'. The 'odd' tens 50 and 70 are expressed as *halv* ('half (way)') to the next ten. This time the ending *-ds* is etymologically the ordinal number, so *treds* (60) is the third score (or 3rd 20), not three score (3x20). It is not completely obvious, because *fir+ds* becomes *fjerds*, but not beyond a linguist's imagination.

You should be able to guess that *halvfems* is 90 based on the pattern of 50 and 70 (you might have expected *halvfemds*, but *halvfems* is close enough), i.e. 'half way to the 5th score'.



P. Jeg kan tælle (2/2)

Some more details and curious facts.

The *s* in 60 and 80 is actually a contraction of *sinds-tyve* i.e. ‘times 20’. The archaic word for 60 is *tre-sinds-tyve*, literally ‘three score’, just like in old English. As mentioned above, the *d* is the ordinal marker, so 70 used to be *halvfjerd-sinds-tyve*.

In modern Danish pronunciation, the *d* in *treds* and *fjerds* is silent, so many people make the mistake of spelling 50 as **halvtres*, as if it were simply ‘half way to tres’, and 70 as **halvfjers*.

Halvfems should be **halvfemts* if the correct spelling of the word for ‘fifth’ was used. The words for 30 and 40 are *tredive* and *fyrre*, disappointingly not **halvandens* and **tos* that the ‘rule’ would predict.

The authorities have tried to introduce a decimal counting system for writing cheques etc (*femti, seksti, syvti, ot(te)ti, niti*), but the number names have not caught other than for official use (e.g. on banknotes: it is a tourist’s nightmare when the note says *femti kronor* but the shopkeeper says *halvtres kronor!*)

Finally, counting in 20s is not so unusual. As already mentioned, it was not uncommon in English even 150 years ago (“Four score and seven years ago our fathers brought forth on this continent a new nation...”), while French still retains the term *quatre-vingts* ‘four twenties’ for 80, except in (some parts of) Switzerland, Belgium, and Canada where the terms *huitante* or *octante* are used. Other European languages with (usually partial) **vigesimal** systems include Welsh, Irish and Basque.

* As is traditional in linguistic description, a star preceding a word indicates that it is a hypothetical word not actually occurring. For example you could say “The plural of *child* ought to be **childs*, but it isn’t.”

Sources:

<http://www.olestig.dk/dansk/numbers.html>

[http://fr.wikipedia.org/wiki/80_\(nombre\)](http://fr.wikipedia.org/wiki/80_(nombre))

<http://en.wikipedia.org/wiki/Vigesimal>

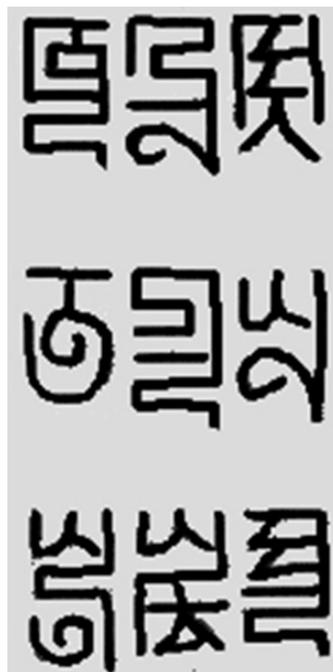


Q. 100 Surnames (1/3)

Q-1

13a.	Kīi	15f.	Kīue	19e.	Tsam
13g.	Maū	15g.	Ma	20b.	Fang
14c.	Sriū	16g.	Yan	20c.	Giū
14e.	Xīang	17a.	Mue	22c.	Tso
15b.	Ngīūan	18b.	Xīa	23d.	Hūa
15d.	Min	18g.	Ling	24c.	‘U

Q-2



Q. 100 Surnames (2/3)

To begin, we can notice that, given the complexity of the glyphs and the frequent recurrence of sub-glyph shapes within them, that the glyphs likely consist of multiple sub-parts.

We can then notice that the diversity of shapes at the top of the glyphs is much greater than the shapes at the bottom: there are a few dozen possible "tops" and only a handle of possible "bottoms". This, when compared to the names in the table, shows that the top shapes of glyphs represent the possible beginnings of names, of which there are many, and the bottoms represent the ends, of which there are few. This imbalance helps to confirm our earlier insight: that sub-glyphs represent sub-parts of names in some systematic manner. It also gives us a hypothesis about the overall writing direction of the text: that it might be top-to-bottom overall as well.

At this point, there are several ways to try to fix what part of the poem these two pages represent. One way is by rhyme scheme: given that the poem is made of 8-name rhyming couplets, there should be a pattern of repeating "bottom" shapes every eight characters. Given the 8x7 shape of the pages, if the text is written horizontally (in either direction), this pattern should occur in vertical lines, and if the text is written vertically (in either direction), this patterns should occur in diagonal lines.

It is easy to see that there are no such vertical-line patterns (especially not at the ends of lines where we would hope them to be). There is, however, an every-eight-characters diagonal rhyme pattern running through the text: from 3rd row/1st column ("3a") upward and rightward to (1c), wrapping around to 7e and then up to 1k, wrapping around again to 7m, then up to 4p.

Going by the bottom shape, the rhyme scheme of the segment pictured here looks to be something like AAAAABBBBAAAAA. There is only one place in the poem this could be: the BBBB glyphs have to represent Kũa, Lau, Fa, and Ma and the rest represent names ending in -ng. This result, however, is still compatible with two writing directions (upward then left-to-right, or downward then right-to-left). The former of these is unlikely given the downward writing direction within glyphs, but it is still at least possible. Looking at the "top" shapes decides it for us: we have an ABCDEFAGCHGIDJ pattern, and this only fits one way with the pattern of name beginnings the text (the downward then right-to-left direction).

This is one way of determining the position and direction of the manuscript relative to the poem; other ways are equally valid and can receive equal points.

(CONTINUED ON THE NEXT PAGE)



Q. 100 Surnames (3/3)

Now that we know which glyphs represent which, we can begin to determine what exactly the sub-glyph shapes represent. If you've come this far, this process will be mostly straightforward. There are four "classes" of glyphs, which I'll call A, B, C, and D:

- A: initial sounds (representing word-initial b, p, dz, kh, y, tr, tsh, etc.)
- B: "on-glides" (ĩ, ũ, etc.), written after initials
- C: vowels (i, u, e, etc.), written after on-glides (if any)
- D: codas (m, ng, ĩ, etc.), written last

Depending on where in a word a sound occurs, a sound like [ng] or [ũ] might have be in a different class and thus have a different shape.

One complication to watch out for is that there is no shape for [a]. If there is no other class C (vowel) glyph in the word, that word's vowel is [a]. You can tell the difference between (say) Hũa and Haũ, even though [a] is not written, by the shape [ũ] gets: the class B or class D shape.

The other complication to note is that when class B [ũ] and [ĩ] co-occur, they are written in 'Phags-pa in an order opposite from what we would expect.



R. One, Two, Tree (1/5)

R-1

- a. *[[ice cream] soda]*
- b. *[[science fiction] writer]*
- c. *[[customer service] representative]*
- d. *[state [chess tournament]]*
- e. *[[Mars Rover] landing]*
- f. *[plastic [water cooler]]*
- g. *[[typeface design] report]*

R-2

[[country song] [platinum album]]

R-3

Default answer: A drama about control freaks (i.e., freaks about control), performed during a space mission (i.e., a mission to space).

Many other answers are possible as long as each of the bracketings is correctly defined. Below are examples of correct answers for each bracketing:

For "*[control freak]*":

- a person who is obsessive about having things his way

For "*[[control freak] show]*":

- a show that is run by control freaks
- a show that contains control freaks (i.e., the actors are control freaks)
- a show that is designed for or intended for control freaks
- a display of behavior by a control freak

For "*[space mission]*":

- a mission into space



R. One, Two, Tree (2/5)

For "[[*space mission*] [[*control freak*] show]"]:

- a control freak show that is broadcast to audiences on space missions
- a control freak show that is set on a space mission
- a control freak show that is about space missions
- a display of behavior by control freaks; the display is witnessed on a space mission

Some examples of incorrect answers:

- the show contains control freaks that are interested in space missions
- the show contains control freaks who run or are on a space mission

These are incorrect because they attach "*space mission*" to "*control freak*" instead of attaching it to "*control freak show*". For these answers to be correct, the bracketing would have to be "[[[*space mission*] [*control freak*]] show]".

R-4

[[[[*family* [*board game*]] [*togetherness effect*]] [*government study*]] *author*].

Although the following might also be defensible:

[[[[*family* [[*board game*] *togetherness*]] *effect*] [*government study*]] *author*]

Or even perhaps:

[[[[[[*family* [*board game*]] *togetherness*] *effect*] [*government study*]] *author*]



R. One, Two, Tree (3/5)

R-5

$$f(5) = 14$$

$$f(6) = 42$$

$$f(7) = 132$$

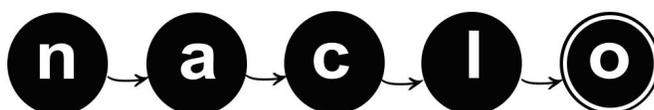
There are $f(5)$ bracketings of *togetherness effect government study author* — whatever $f(5)$ turns out to be! Similarly there are $f(3)$ (= 2) bracketings of *family board game*. So you have to list $f(3) \cdot f(5)$ bracketings that split the 8-word sequence into 3 words + 5 words like this. But the full list for $f(8)$ must also consider other splits, such as 1 word + 7 words. The general principle is that

$$f(n) = \sum_{k=1}^{n-1} f(k) \cdot f(n-k) \text{ for any } n > 1$$

You can therefore compute each line in the table from the previous lines. By the way, the resulting sequence of numbers is called the Catalan numbers: you can look it up.

R-6

To get $[[big\ fluffy]\ pancake]$, we'd need adjective + adjective = adjective. To get $[[samurai\ short]\ sword]$, we'd need noun + adjective = adjective or noun + adjective = noun.



R. One, Two, Tree (4/5)

R-7

a. There are only 3 bracketings (fewer than $f(4)=5$ because the rules from R-6 are “missing”):

[roasted [red [potato pancake]]] - a roasted red pancake made of potatoes

[roasted [[red potato] pancake]] - a roasted pancake made of red potatoes

[[roasted [red potato]] pancake] - a pancake made of roasted red potatoes

Note that the 4th logical possibility, a red pancake made of roasted potatoes, is not consistent with this word order: you'd have to call it a *red roasted potato pancake*.

b. There are 7 bracketings (fewer than $f(5)=14$):

[[crazy monkey] [[cheap cider] house]] - the house of crazy monkeys serves cider that is cheap

[[crazy monkey] [cheap [cider house]]] - the house of crazy monkeys that serves cider is cheap

[crazy [monkey [[cheap cider] house]]] - the crazy house of monkeys serves cider that is cheap

[crazy [monkey [cheap [cider house]]]] - the crazy house of monkeys that serves cider is cheap

[crazy [[monkey [cheap cider]] house]] - the crazy house serves cheap cider that's for monkeys

[[[crazy monkey] [cheap cider]] house] - the house serves cheap cider that's for crazy monkeys

[[crazy [monkey [cheap cider]]] house] - the house serves crazy, cheap cider that's for monkeys

Again, there are logical possibilities that are not consistent with the word order, such as a house of monkeys that serves crazy, cheap cider.

R-8

a. 0

b. 1

c. 5. This generalizes R-7a. There are 5 ways to divide *Adj Adj Adj Adj* into an initial group that modifies the 2nd noun and a final group that modifies the 1st noun. Groups may be empty.

d. 14. First suppose that the nouns are bracketed as $[[Noun\ Noun]\ Noun]$. Then there are 10 ways to divide *Adj Adj Adj* into 3 groups which will respectively modify the 3rd, 2nd, and 1st noun (There are $4+3+2+1=10$ ways to place two vertical dividers into this sequence. If the left divider falls before the first *Adj*, there are 4 positions for the right divider. If it falls before the second *Adj*, there are 3 positions for the right divider; and so on.) Alternatively, suppose that the nouns are bracketed as $[Noun[Noun\ Noun]]$. Then no adjective can modify the 2nd noun (it can only modify the whole $[Noun\ Noun]$ compound), so then we divide *Adj Adj Adj* into only 2 groups as before; there are 4 ways to do this.



R. One, Two, Tree (5/5)

R-9

- a. 1
- b. 2
- c. 3
- d. 7
- e. 12
- f. 30

Let $g(n)$ be the number of bracketings of an alternating n -word sequence ending in *Noun*. Clearly, $g(1)=1$. To find $g(n)$, we can proceed as in R-5 and consider the ways of splitting the sequence into two shorter sequences, whose bracketings we count by applying g recursively. However, we have to leave out the splits where the first sequence ends in an adjective (i.e., where the second sequence has odd length), unless it's a single adjective (i.e., the first sequence has length 1). Thus, for odd $n>1$,

$$g(n) = g(n-2)g(2) + g(n-4)g(4) + \dots + g(1)g(n-1),$$

while for even $n>1$,

$$g(n) = g(n-2)g(2) + g(n-4)g(4) + \dots + g(2)g(n-2) + 1 \cdot g(n-1)$$

We can therefore compute

$$\begin{aligned} g(2) &= g(1)g(1) &&= 1 \\ g(3) &= g(1)g(2) &&= 1 \\ g(4) &= g(2)g(2) + 1 \cdot g(3) &&= 2 \\ g(5) &= g(3)g(2) + g(1)g(4) &&= 3 \\ g(6) &= g(4)g(2) + g(2)g(4) + 1 \cdot g(5) &&= 7 \\ g(7) &= g(5)g(2) + g(3)g(4) + g(1)g(6) &&= 12 \\ g(8) &= g(6)g(2) + g(4)g(4) + g(2)g(6) + 1 \cdot g(7) &&= 30 \end{aligned}$$

Notice that these formulas are much faster than *listing* all the bracketings, which is important since $f(25) = 1,289,904,147,324$ and even $g(25) = 50,067,108$. The approach here can be generalized into a single powerful technique for rapidly counting the number of bracketings of *any* given word sequence.

Closely related algorithms are used to rapidly find the most likely bracketing or “parse” of a given sentence so that a computer can understand or translate it.

