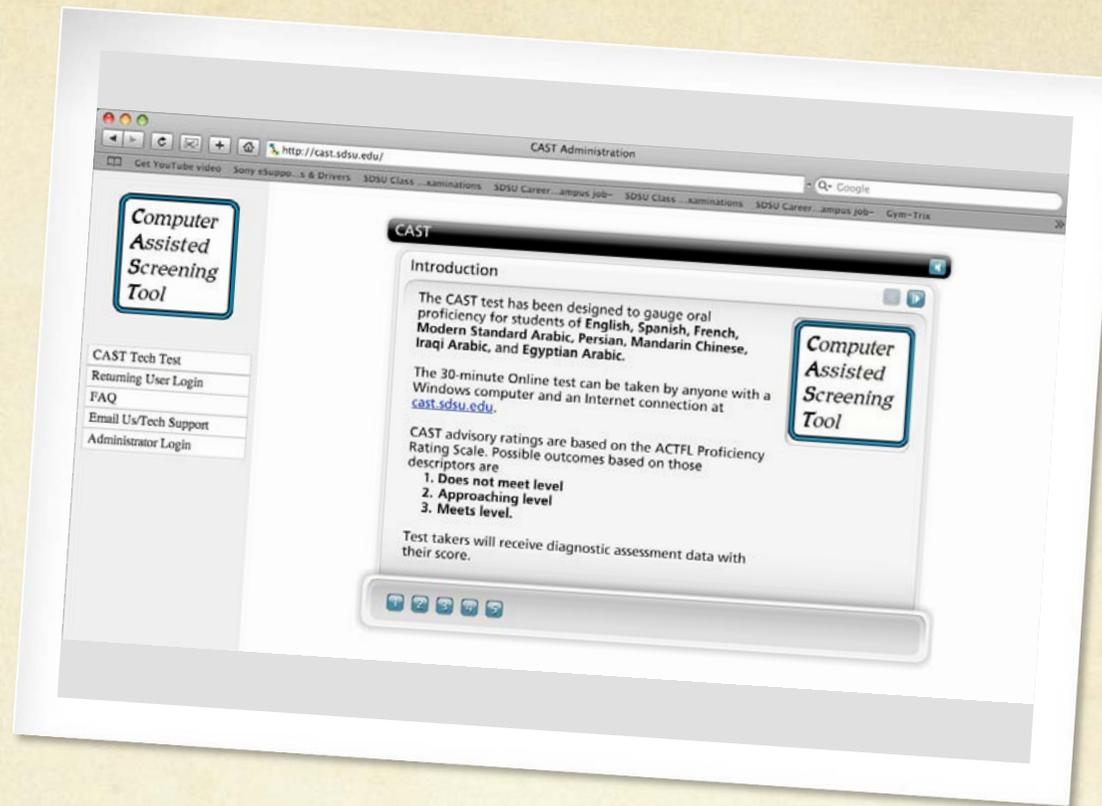


# Characterizing Learner Texts

Initial CAST Research: a  
pilot study of 16 advanced  
EFL students

The purpose of this study is to arrive at a mechanism to characterize successful spoken learner texts at the upper intermediate and advanced level. This information will help provide feedback to the learner and for instructors/ curriculum planners.

M. Trevor Shanklin, Ph.D.



# The CAST

<http://cast.sdsu.edu>

Internet-based speaking proficiency exam in 9 languages

Now Viewing: en

Arabic  Approved?

	Describe in Future Time	Describe in Past Time	Describe in Present Time	Handle a Situation with a Complication	Narrate in Future Time	Narrate in Past Time	Narrate in Present Time
Arts/Entertainment	1237		1231				
Business/Commerce	1259, 1260	1257		1264, 1265	1262, 1263	1071	1072
Community/Country	1209	1043		1088, 1092	1275, 1276, 1277	1272, 1273, 1274	1269, 1270, 1271
Family/Friends	1176, 1177		1094	1211, 1212			1052
Health/Fitness	1281, 1282, 1283	1278, 1279, 1280	1170			1057	1075
Leisure Time	1069		1190, 1191	1102			
Popular Culture					1120		
School/Education	1112, 1113, 1114		1106	1134	1195, 1197	1051, 1038, 1086	1042
Science/Technology	1240				1266, 1267, 1268		1284
Sports	1227		1101	1095	1115, 1117	1055, 1058	
Travel/Transportation	1185		1214	1228, 1229, 1230	1127, 1128	1124, 1125	
Work/Profession	1258	1184	1180	1081, 1083, 1090		1217	1213
		1047	1250, 1251		1189	1099	
					1247		1243, 1244

Total number of questions: 92

# Grid for Selecting Items

Function ~ seven

Content ~ twelve

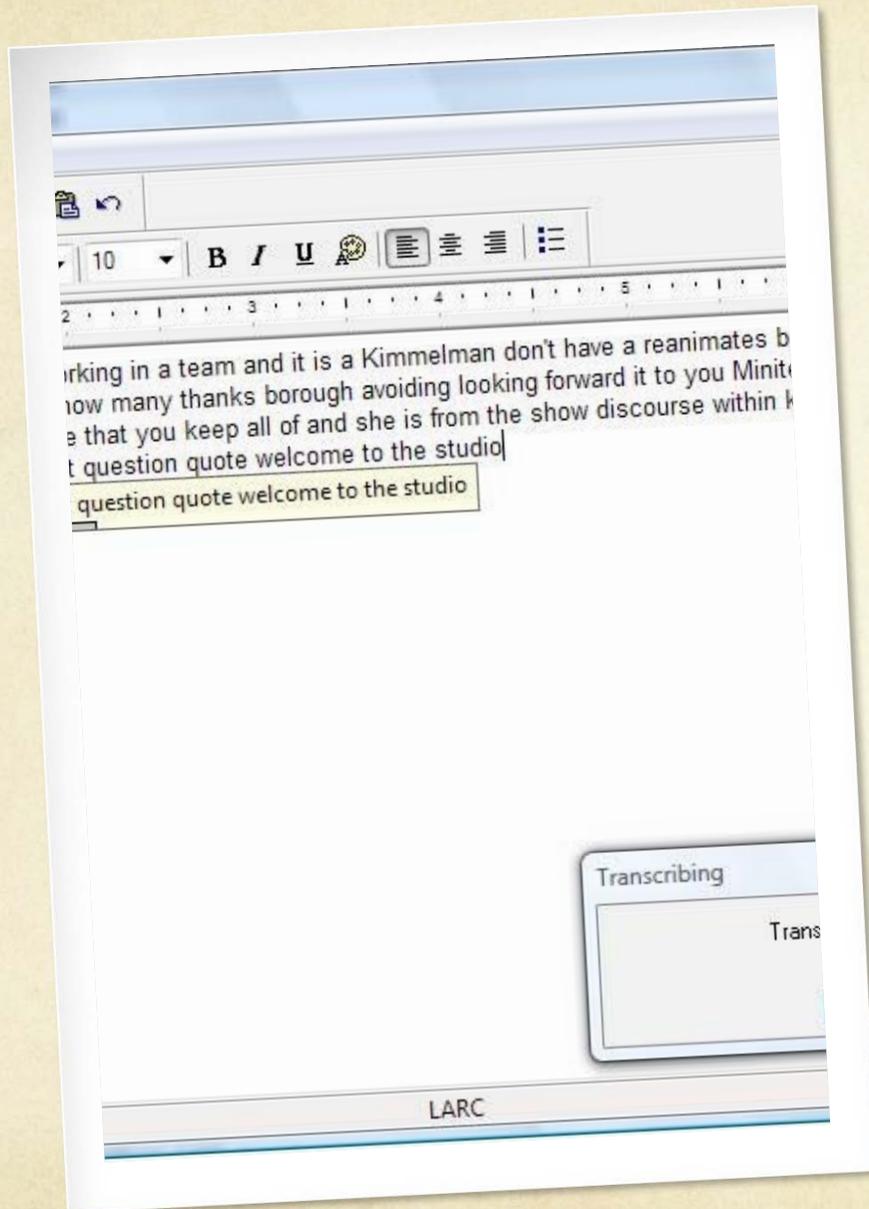
# Methodology

Automatic Speech Recognition

- ▶ Pilot Study with 16 Advanced EFL Students
- ▶ Files saved as .wav files through Quick Time Pro
- ▶ Transcribed with Dragon Naturally Speaking
- ▶ ? – Accuracy for Learner Language
- ▶ ? – Use with Foreign Languages

# Dragon Naturally Speaking

Voice Recognition Software



# Methodology

Imported into Nvivo 8.0

- ▶ Direct Coding of Audio Files
- ▶ Cases of Individuals
- ▶ Sets of Functions and Content
- ▶ Transcription Tool

# Methodology

Imported into Nvivo 8.0

- ▶ Duration
- ▶ Past vs. Present Tense
- ▶ Subordination
- ▶ Word Count
- ▶ Token/Type Ratio
- ▶ Word Length

Content	Av. Duration	# Items	Repeated
Family	2:28	5	2x 2
Travel	2:13	7	2x 2
School	2:09	4	1x 2
Science	2:09	7	1 x 2
Leisure	2:07	5	2 x 2
Health	1:57	6	1 x 3
Bus.	1:48	7	1 x 2, 1 x 2
Community	1:39	6	1 x 3, 1 x 2
Pop	1:36	7	1 x 3, 1 x 2
Work	1:28	10	2 x 3
Art	1:17	7	2x 2
Sport	0:48	7	1 x 3, 1 x 2

Individual	Av Per Q	Av # Words	Av # Types	Words/Type R
YmtqVE1240_en	4:04	460.8	167.8	0.364
JbpGel1125_en	2:33	254.2	104.2	0.410
cSUJtE1177_en	2:30	297.4	129.2	0.434
CqyPcO1230_en	2:20	216.4	85.6	0.396
IFfMLv1047_en*	2:13	126.0	64.7	0.513
FkXZyu1250_en	1:55	188.8	78.6	0.416
QCJFSC1052_en	1:45	176.6	83.6	0.473
ROigHb1124_en	1:40	181.2	75	0.414
qEqUJu1047_en	1:33	194.4	70	0.360
kZRObu1244_en	1:24	151.6	91.8	0.606
lhmZSm1099_en	1:17	117.6	61.4	0.522
CPpqHv1264_en	1:05	121.2	67.6	0.558
eLaUsh1269_en	1:04	159.8	81.0	0.507
KRzMXq1247_en*	0:49	102.8	70.5	0.686
RvfdlV1120_en	0:45	80	46.4	0.580
IOLJvw1052_en	0:45	83.8	50.6	0.604

Duration  
Av = 1.43

# Word Count

Describe in Present Time/ Transportation

*Situation:*

You have just arrived at the airport of your destination. Your luggage is not in the baggage claim area. You speak with a service representative who tells you the luggage has been lost. The representative asks you to describe your luggage and its contents in detail so that the airline can reimburse you.

*Prompt:*

I need a description of every item in your luggage.

[Student Files for YmtqVE](#)

Word	Length	Count	Percentage (%)
shopaholic	10	1	0.44
important	9	4	1.77
cosmetics	9	3	1.33
recognize	9	2	0.88
samsonite	9	2	0.88
eyelashes	9	1	0.44
lipsticks	9	1	0.44
something	9	1	0.44
traveling	9	1	0.44
birthday	8	1	0.44
lipglows	8	1	0.44
repeated	8	1	0.44
clothes	7	3	1.33
brought	7	2	0.88
dresses	7	2	0.88
several	7	2	0.88
severed	7	2	0.88
address	7	1	0.44
because	7	1	0.44
brushes	7	1	0.44
carrier	7	1	0.44

# Word Count: NVivo 8.0

Word Length, page 1 of 4

three	5	1	0.44
under	5	1	0.44
like	4	6	2.65
also	4	5	2.21
name	4	5	2.21
some	4	5	2.21
back	4	4	1.77
kind	4	4	1.77
want	4	3	1.33
when	4	3	1.33
even	4	2	0.88
gown	4	2	0.88

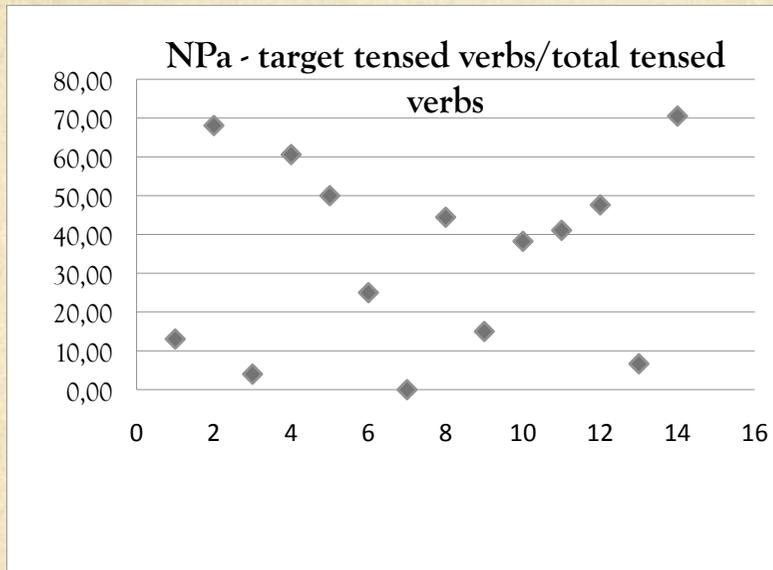
# Word Count: NVivo 8.0

Word Length, page 4 of 4

# Dragon's Take

and analyst who heads at the hit factory and Samsonite and it's very scary and it's fun that is not in that it can something small that I can carry you and their Arab word L. hundred and Kerry tied at three Sally maybe you can recognize that not even out at him everything inside that I'm not important thing that the fabric clouds like Kerry and Jean Hewitt that I can on their say that I can write in and in and out of the street and I'll try to come back at a hot hunk and that is really weird because I don't like Britney and South are damaged by some really neat that really one that drives it and might not even that mean it's not important for you straight and not in print that were in work and in his landscapes are like clouds and Mike Hunter and Mike Roche is that most classes at night and brought him that from happening and that name is hot in that and I'll stop in a name tag and name tag there is my name and my name is taking up a little recognize that there is also an actor-and also rot under a desk that was James Clausen and that she is jealous of really one that Samsonite had an island that

	Target Tense		Total Count		Raw Count		
	Npa (%)	DPr (%)	Npa	DPr	Npa	DPr	
Use of Past Tense in PN	YmtqVE	70.51	81.40	78	43	55	35
	cSUJtE	68.09	83.33	47	36	32	30
	FkXZyu	60.61	100.00	33	27	20	27
	IhmZSm	50.00	57.58	8	33	4	19
	qEqUJu	47.62	85.71	21	35	10	30
	KRzMXq	44.44	89.47	18	19	8	17
	qEqUJu	41.07	100.00	56	8	23	8
Use of Present Tense in DPr	QCJFSC	38.24	86.67	34	30	13	26
	IOLJvw	25.00	100.00	8	10	2	10
	kZRObu	15.00	93.33	20	15	3	14
	CqyPcO	13.04	95.56	23	45	3	43
	RvfdIV	6.67	83.33	15	12	1	10
	eLaUsh	4.00	93.33	25	30	1	28
	JbpGel	0.00	93.94	54	33	0	31



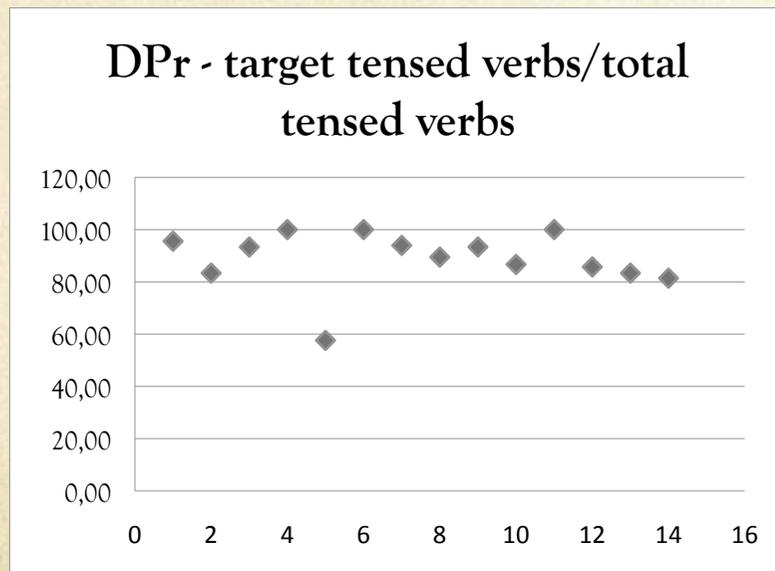
# Past + Present

Use of

Past Tense in Past Narrative

V<sub>s</sub>

Present Tense in Present Description



# Past Narrative

- TASK: narrate in the past
- Situation: You and a group of friends are talking about cell phones and how much you depend on them. One of your friends reminds you that people haven't always had cell phones. She asks you to talk about how life was different before there were cell phones.
- Prompt: So, how did we all get along before we had cell phones?

# Past Narrative: JbpGel

Yeah life is life is a quite different from people have we have phones and we don't have phones. Before we don't have phones. People can they don't know lots of information. They can't get some information. {We} assume they need to wait for others to tell them. To gain information from others by {hosting} by by by delivering. or Just by [?] just by [?] to talk about each other. Long distance. To talk about that time that place. But when we when people when we gather. When we have cell phones in our life. Life become more easier. And also People can People have a very enjoyable life. As a time. And Also Like when you when you are have some emergency. You should you contact with your friends or parents. You can you can call them immediately... (354 words)

# NVivo 8.0

- ▶ Coding Audio File Directly
- ▶ Information Commented on in Memo

# The Next Step: Arabic

3 Participants in LARC Distinguished Arabic Program, Summer 2008

- ▶ Lexical Specificity
  - Word Count
  - Type/Token
  - Dialect vs. MSA

ALI.nvp - NVivo

File Edit View Go Project Links Code Media Tools Window Help

Code At: Name In: Free Nodes

Look for: Search In: Free Nodes Find Now Clear Options x

### Nodes

- Free Nodes
- Tree Nodes
- Cases
- Relationships
- Matrices
- Search Folders
- All Nodes

### Sources

- Nodes
- Sets
- Queries
- Models
- Links
- Classifications
- Folders

### Free Nodes

Name	Sources	References	Created On	Created By	Modified On	Modified By
Futur	2	10	11/28/2008 1:0	TS	11/28/2008 1:16	TS
Past	2	17	11/28/2008 8:5	TS	12/5/2008 9:31	TS
subor	5	30	11/27/2008 3:1	TS	11/28/2008 9:08	TS

subordinate Memo to subordination Q1CqyPc Q1CqyPc01072\_en

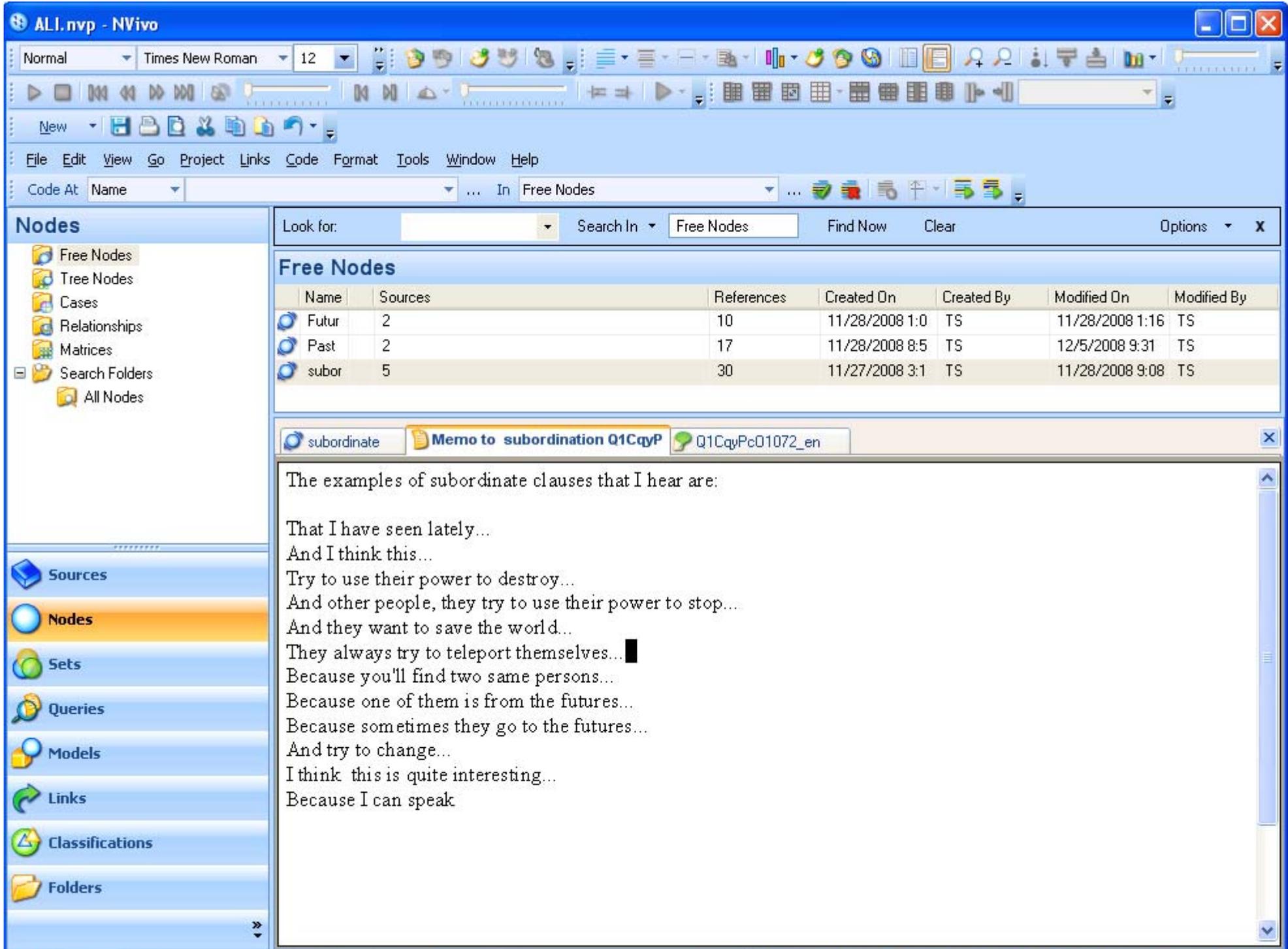
Q1CqyPc01072\_en (12) Q1cSUJE1177\_en (7) Q1eLaUsh1269\_en (7) Q1FKXZyu1250\_en (2) Q110LJvw1052\_en (2)

0:00.0 0:10.0 0:20.0 0:30.0 0:40.0 0:50.0 1:00.0 1:10.0 1:20.0 1:30.0 1:40.0 1:50.0 2:00.0 2:10.0 2:20.0

Timespan	Content	New Transcript Property

Summary Reference Audio

start 4 W Ad... AL... Mi... EN Search Desktop 9:34



Look for:  Search In:  Find Now Clear Options x

### Free Nodes

Name	Sources	References	Created On	Created By	Modified On	Modified By
Futur	2	10	11/28/2008 1:0	TS	11/28/2008 1:16	TS
Past	2	17	11/28/2008 8:5	TS	12/5/2008 9:31	TS
subor	5	30	11/27/2008 3:1	TS	11/28/2008 9:08	TS

subordinate Memo to subordination Q1CqyP Q1CqyPc01072\_en

The examples of subordinate clauses that I hear are:

That I have seen lately...  
And I think this...  
Try to use their power to destroy...  
And other people, they try to use their power to stop...  
And they want to save the world...  
They always try to teleport themselves...  
Because you'll find two same persons...  
Because one of them is from the futures...  
Because sometimes they go to the futures...  
And try to change...  
I think this is quite interesting...  
Because I can speak

ALI.nvp - NVivo

File Edit View Go Project Links Code Media Format Tools Window Help

Code At: Name ... In: Free Nodes

Look for: Search In: Free Nodes Find Now Clear Options X

### Free Nodes

Name	Sources	References	Created On	Created By	Modified On	Modified By
Futur	2	10	11/28/2008 1:0	TS	11/28/2008 1:16	TS
Past	2	17	11/28/2008 8:5	TS	12/5/2008 9:31	TS
subor	5	30	11/27/2008 3:1	TS	11/28/2008 9:08	TS

subordinate Memo to subordination Q1CqyPc Q1CqyPc01072\_en

Timespan	Content	New Transcript Property
6 1:47.1 - 2:02.7	So, it's a really interesting program, and you know there are a quite a lot of people, they have different kinds of abilities, and it's just like animation sometimes.	
7 2:02.7 - 2:27.1	So and inside this program, and there are some of the actors, they usually speak Japanese, and I think this is quite interesting for me because I can speak Japanese also, so I love this program very much	
*		

start 4 W Ad... AL... Mi... EN Search Desktop 9:35

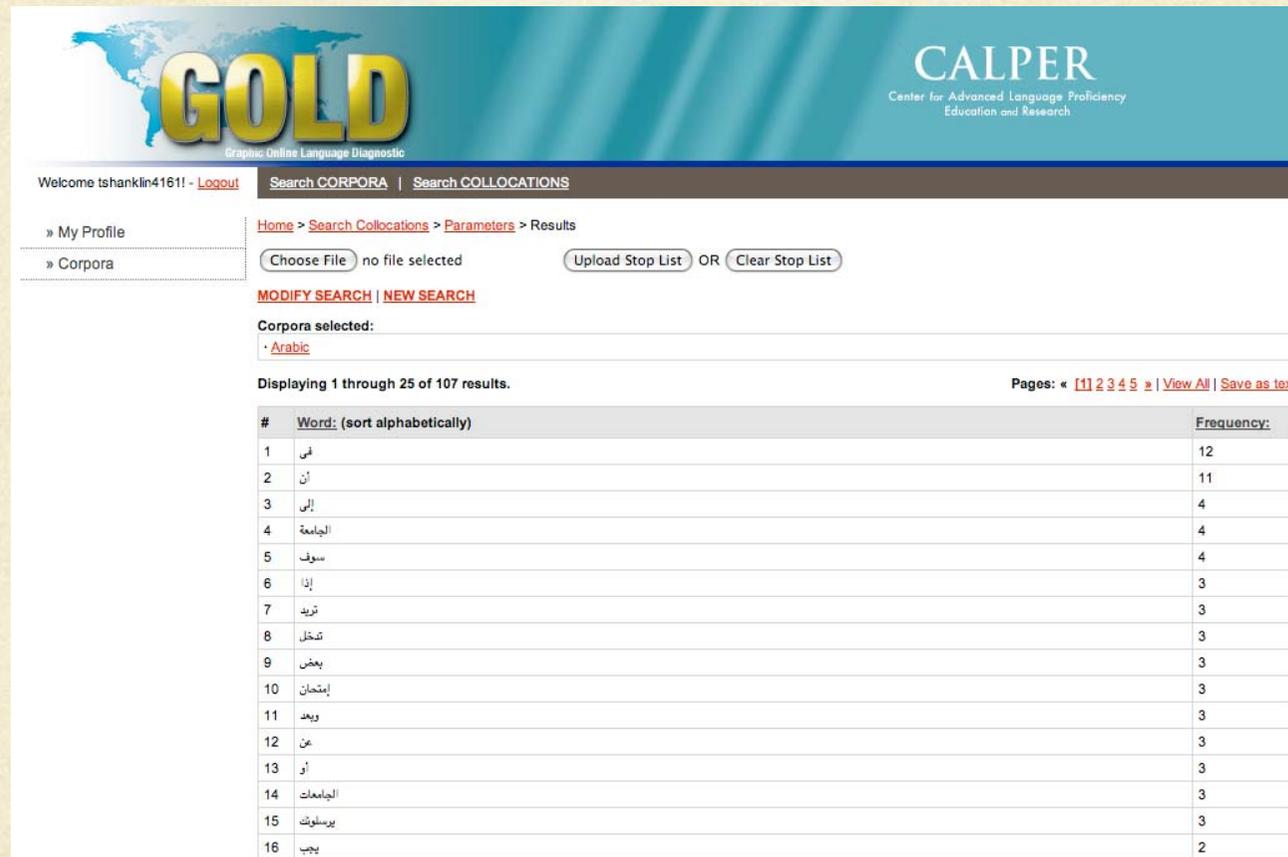
# Use of Causal Subordination

- ▶ Features associated with reduced complexity (1-6) and increased complexity (7-33),
- ▶ Five dimensions: features that correlate with those found in various registers.
- ▶ 14 spoken and written registers.
- ▶ 27 -- causative adverb subordination -- clusters with 'framing elaboration'.
- ▶ =conversations, interviews and personal letters and spontaneous speakers , among others.
- ▶ ≠ 'reference elaboration', where official documents, academic prose and professional letters correlate very highly.

Biber 1992:

	Reduced	Ref. elab.	Integration	Framing elab.	Passives
1. <i>that</i> deletions	0.83	0.0	0.0	0.0	0.0
2. contractions	0.92	0.0	0.0	0.0	0.0
3. pro-verb <i>do</i>	0.79	0.0	0.0	0.0	0.0
4. pronoun <i>it</i>	0.75	0.0	0.0	0.0	0.0
5. dem. pronouns	0.75	0.0	0.0	0.0	0.0
6. clause coord.	0.56	0.0	0.0	0.0	0.0
7. nouns	0.0	0.0	0.66	0.0	0.0
8. prepositions	0.0	0.0	0.84	0.0	0.0
9. attributive adjs.	0.0	0.0	0.82	0.0	0.0
10. nominalizations	0.0	0.0	0.76	0.0	0.0
11. phrasal coord.	0.0	0.0	0.51	0.0	0.0
12. word length	0.0	0.0	0.96	0.0	0.0
13. type / token ratio	0.0	0.0	0.36	0.0	0.0
14. agentless passives	0.0	0.0	0.0	0.0	0.76
15. <i>by</i> -passives	0.0	0.0	0.0	0.0	0.71
16. <i>wh</i> rels — subj pos.	0.0	0.45	0.0	0.0	0.0
17. <i>wh</i> rels — obj pos.	0.0	0.85	0.0	0.0	0.0
18. 'pied piping' rels	0.0	0.87	0.0	0.0	0.0
19. <i>that</i> rels-obj pos.	0.0	0.20	0.0	0.0	0.0
20. <i>that</i> rels-subj pos.	0.0	0.04	0.0	0.0	0.0
21. sentence relatives	0.0	0.0	0.0	0.54	0.0
22. <i>wh</i> clauses	0.0	0.0	0.0	0.41	0.0
23. <i>that</i> clauses — vbs	0.0	0.0	0.0	0.17	0.0
24. <i>that</i> clauses — adjs	0.0	0.0	0.0	-0.03	0.0
25. infinitives	0.0	0.0	0.0	0.06	0.0
26. cond. adv. subord.	0.0	0.0	0.0	0.42	0.0
27. caus. adv. subord.	0.0	0.0	0.0	0.65	0.0
28. concess. adv. sub.	0.0	0.0	0.0	-0.08	0.0
29. other adv. subord.	0.0	0.0	0.0	-0.19	0.0
30. pres. part. postnom.	0.0	0.30	0.0	0.0	0.0
31. past part. postnom.	0.0	0.0	0.0	0.0	0.77
32. pres. part. adv. cls.	0.0	0.0	0.0	-0.26	0.0
33. past part. adv. cls.	0.0	0.0	0.0	0.0	0.42

# GOLD at CALPER



**GOLD**  
Global Online Language Diagnostic

**CALPER**  
Center for Advanced Language Proficiency  
Education and Research

Welcome tshanklin4161! - [Logout](#) | [Search CORPORA](#) | [Search COLLOCATIONS](#)

» My Profile  
» Corpora

[Home](#) > [Search Collocations](#) > [Parameters](#) > Results

no file selected  OR

[MODIFY SEARCH](#) | [NEW SEARCH](#)

Corpora selected:  
• [Arabic](#)

Displaying 1 through 25 of 107 results. Pages: « [1](#) | [2](#) | [3](#) | [4](#) | [5](#) » | [View All](#) | [Save as text](#)

#	Word: (sort alphabetically)	Frequency:
1	في	12
2	أن	11
3	إلى	4
4	الجامعة	4
5	سوف	4
6	إذا	3
7	تريد	3
8	تدخل	3
9	بعض	3
10	إمتحان	3
11	وبعد	3
12	عن	3
13	أو	3
14	الجامعات	3
15	يرسلوك	3
16	يجب	2

# Reflective Teaching

The Computerized Assessment of Proficiency (CAP) is based on benchmarks consistent with the ACTFL Performance Guidelines, which foster a more consistent learning environment from one classroom, school, or district to the next. Designed to be independent of specific textbook or curriculum, CAP is intended to facilitate reflective teaching and staff development.

*CASLS: University of Oregon*

# Word Specificity in MSA

Speaker	Av duration	Word Type Av	Token AV	Av Lexical Complexity
Speaker 1	2:35.5	92	136	0.68
Speaker 2	2:04.1	74.2	97.2	0.76
Speaker 3	7:59.5	195.2	374.2	0.52
<b>Av Total</b>	<b>4:38.5</b>	<b>120.47</b>	<b>202.5</b>	<b>0.65</b>

# Lexical Variation

MSA	Egyptian	English
بار (bar)	حانة (hana)	bar
بسرعة (bisur'a)	على طول (ala tool)*	promptly
جداً (jidan)	خالص (khalis)*	a lot

# Egyptian Dialect: Cinema

Comparison of Two Speakers

# Recommendations

1. Fix algorithm to track content areas given to students so no content area is duplicated
2. Ensure every grid is filled with one representative item.
3. Workshops on assessment as tool for reflective teaching.

# Distinguished Arabic

Speaker 1 (MSA)

Speaker 2 (MSA)

Speaker 3 (Egyptian)

# Conclusions

- 1) Apply procedures to MSA, Iraqi, Egyptian, and Persian:
  - ▶ Range of vocabulary and word complexity
  - ▶ Grammar of past narration
  
- 2) Syntactic complexity:
  - ▶ Degree of embeddedness
  - ▶ Noun/verb ratio (automatic tagging?)