TransformELT

EMaDA & RIA Interim Progress Sharing Event

10/12 January 2023

Supported by

BRITISH COUNCIL

www.britishcouncil.org

TransformELT

EMaDA & RIA Interim Progress Sharing Event

Corpus of Native-speaker Youth English

10/12 January 2023

Supported by

BRITISH COUNCIL

TransformELT

EMaDA & RIA Interim Progress Sharing Event

James THOMAS

versatile.pub@gmail.com

Alan Pulverness

Project manager

10/12 January 2023

Prof. Gong
Basic Education Curriculum and Teaching Material Research Center.

Fraser Bewick
British Council

Supported by

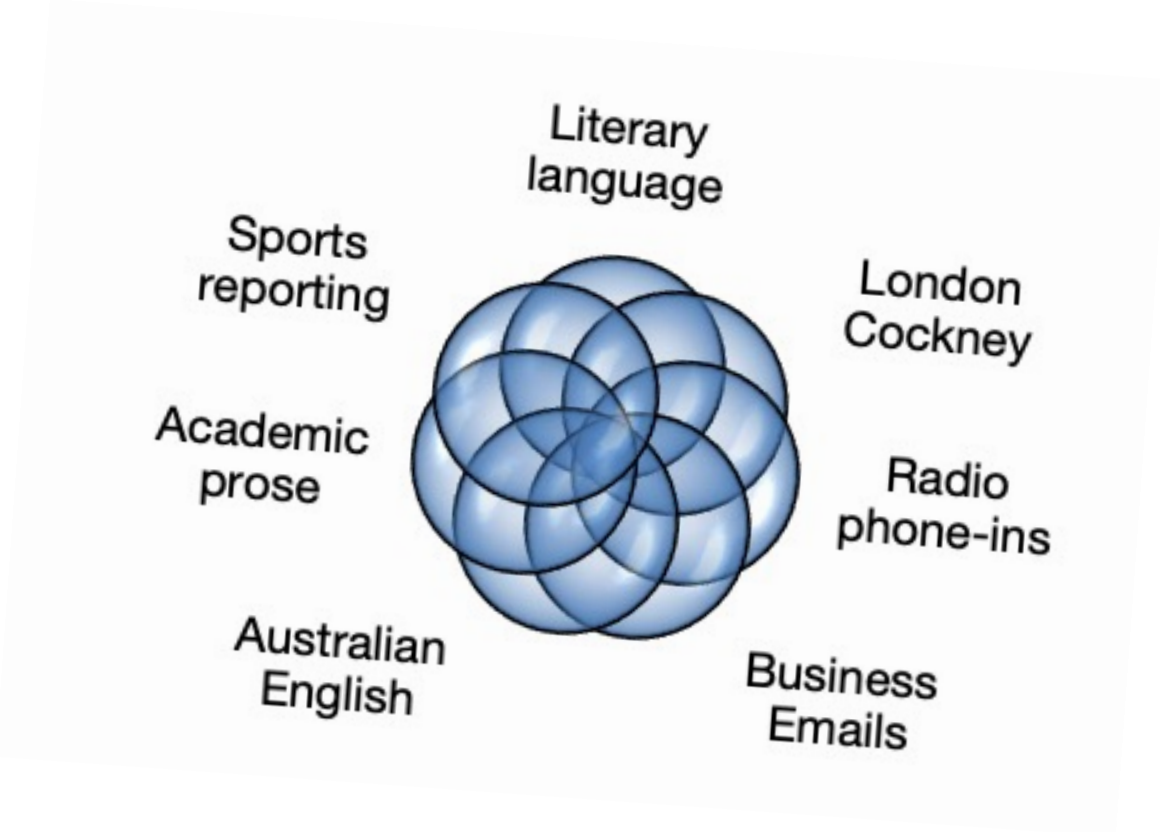
BRITISH COUNCIL

Introduction

- 80% of the 15 million English teachers worldwide are NNESTs. (Freeman et.al., 2015)
- Localising resources
  - ➢ Culture
  - ➢ Motivation
  - ➢ Relevance
  - ➢ Representation
- Young learners
- Learning English
  - ➢ Preparation for the future
  - ➢ A large part of their present
  - ➢ Engage in the cultures of the language
  - ➢ Strong foundation in General English

Creating compatible resources

- An unhealthy, unhelpful, unrealistic and distracting focus on exotic aspects of young people's vocabulary
- Interest in slang, idiomatic and figurative language, informal speaking and writing, taboo words.
- Core vocabulary, core usage.
- *hope, decide, system, ask, answer, after, always, flower, hurt, know, phone, picture, place, read, say, school, much, short, hear, back, beautiful, careful, caring, hungry, middle, tidy, trend, young.*
- English for specific purposes:
- subject terminology uses core vocabulary in core structures.
- CLIL, EMI, Content-based teaching, ESP …



Venn diagram of some genres and registers

Research questions

do vs. can

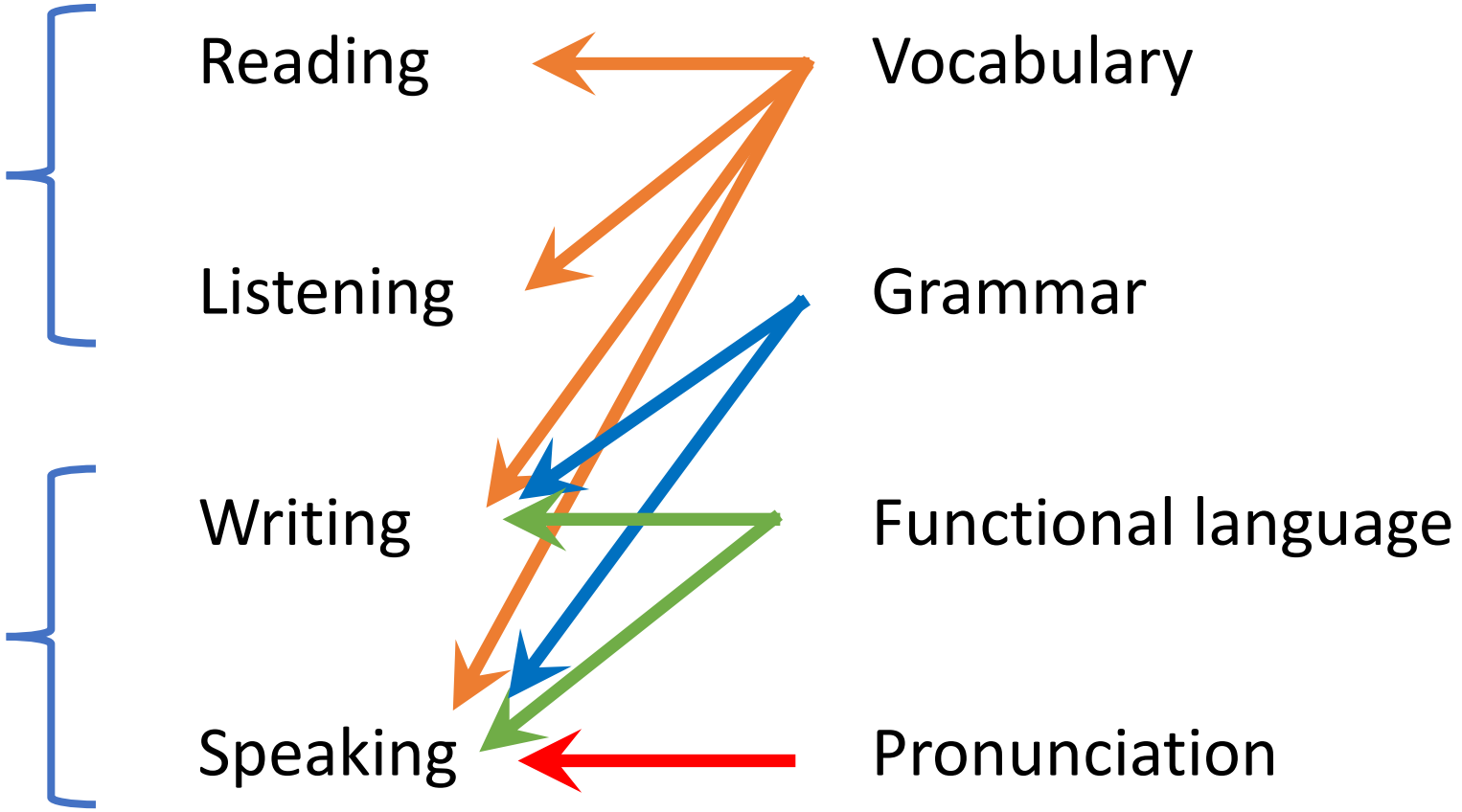
- How do young native speakers use the words on the CNEC wordlists?
- What words and phrases do young native speakers commonly use that are not on the lists?
- How do you know?
- What use is this information?

Patterns of normal usage

The four skills and the four systems

Receptive Input

Productive Output



How people use words

Receptive purposes: **input**

Reading and listening

- Meaning of the word / phrase
  - (Connotations)
  - (Contextual variation)
- Input informs output
- Learn and acquire the norms of the language

Productive purposes: **output**

Writing and speaking

1. Choose the right word
2. Co-select the words that co-create the intended meaning
3. Combine them into a message chunk
4. Combine message chunks into text

How people use words

Recipe for creating a message unit

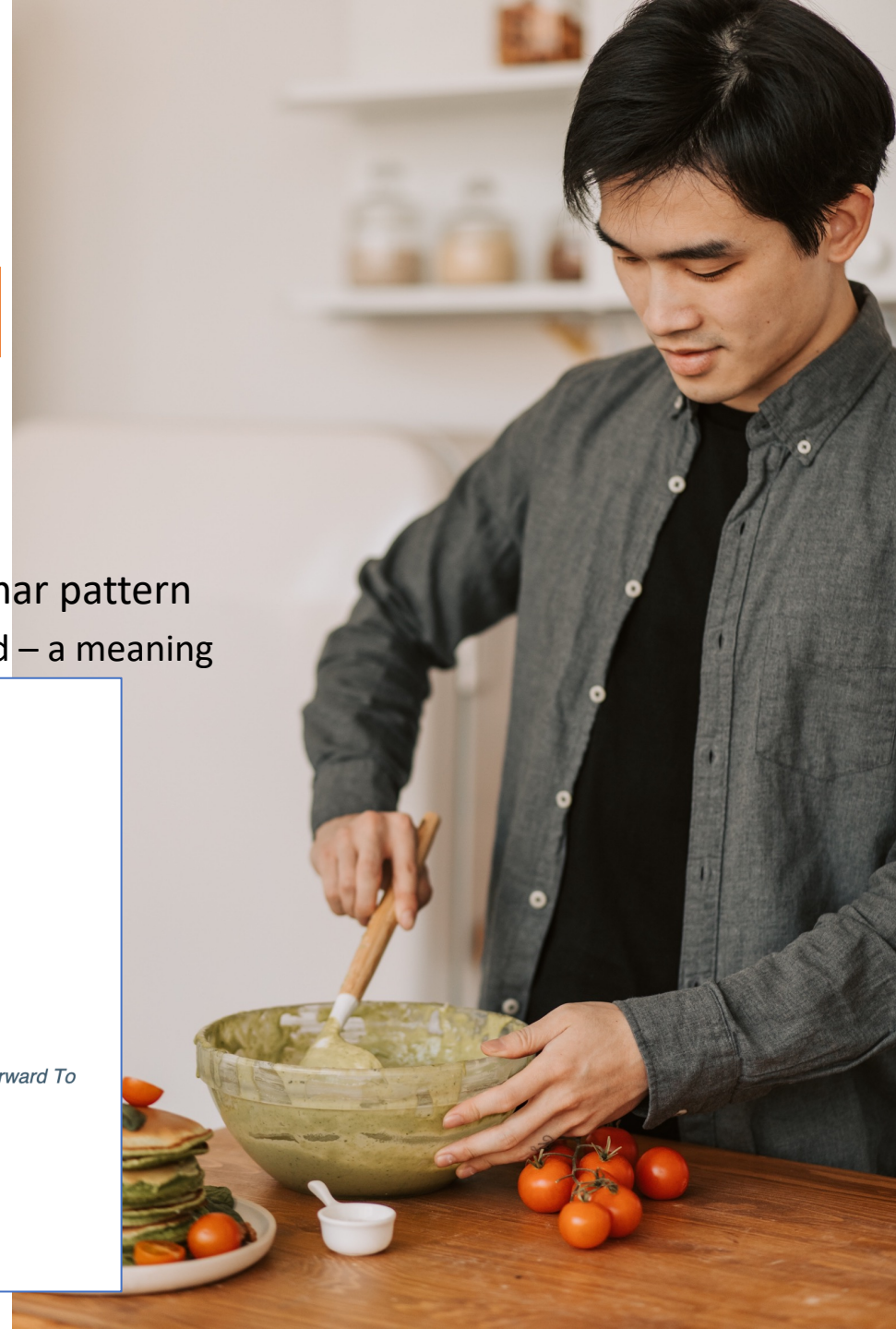
Ingredients

- Key word – usually a noun
- Other content words – collocation
- Function words – colligation
- Co-selection
- Collocation: limited choice
- Colligation: no choice

Method

- Combine according to grammar pattern
- Each grammar pattern of a word – a meaning

N	adjective face	*Deck*
N	poss face	*Face*
N	adjective face	*Hair And Heart*
N	to noun (my)	*Other*
N	adjective face	*Other*
N	face of noun	*Rim*
N	noun face	*Surface*
N	in face (the)	*Trouble*
Vb	face noun with noun	*Beset*
Vb	face noun with noun	*Bore*
Vb	face -ing (not)	*Dread And Look Forward To*
Vb	face noun	*Face*
Vb	face prep/adv, face adv/prep	*Face*
Vb	face noun	*Hear*
Vb	face noun (cannot)	*Hear*
Vb	face noun	*Other*



Corpus of Native-speaker Youth English

Tokens	54,130,818
Words	44,287,785
Sentences	5,855,792
Documents	5,533

How do you know how people use words?

Input corpus

- To explore large samples of English that have been written for young native speakers of English.

SKETCH ENGINE

Output corpus

- To explore large samples of English produced by young native speakers of English.
  - ➢ Relevant sections of CHILDES
  - ➢ Relevant sections of the BNC Spoken corpus
  - ➢ Samples of written English by young people

Corpus of Native-speaker Youth English

What we can extract from the corpus

How words are used (patterns of normal usage)

1. CNEC words in all POS
2. Grammar patterns
3. Bundles
4. Phrases
5. Collocations:
   (a) contiguous
   (b) gramrels
6. Sentences

1. Conversion e.g., *a face, to face.*
2. Extended colligation: e.g., *Verb noun with noun*
3. A group of consecutive words that occur in texts e.g., *in the face of, the face of it*
4. A group of words which has a holistic meaning, e.g. *the look on her face, turn to face me*
5. Next slide
6. Illustrative sentences of the words in these combinations.

Collocations

Contiguous

the words are adjacent in the language

e.g., adj + noun, adv + verb.

Gramrels ***(grammatical relationships)***

Range of ±4.

e.g. objects of verb, modifier with noun.

face	be	*noun*	*verb*	822
face	value	*noun*	*noun*	163
face	up	*verb*	*adverb*	152
face	have	*noun*	*verb*	107
face	look	*noun*	*verb*	81
face	be	*verb*	*verb*	80
face	down	*verb*	*adverb*	79
face	off	*verb*	*adverb*	73
face	turn	*noun*	*verb*	56
face	mask	*noun*	*noun*	49
face	shape	*noun*	*noun*	47
face	page	*verb*	*noun*	43
face	get	*noun*	*verb*	41
face	difficulty	*verb*	*noun*	36
face	become	*noun*	*verb*	36
face	challenge	*verb*	*noun*	34
face	show	*noun*	*verb*	32

TransformELT

VOCABULARY LISTS
in the CNEC

言

Through the eyes of
CONYE

From corpus to database

How words are used ➔ patterns of normal usage

1. CNEC words in all POS
2. Grammar patterns
3. Bundles
4. Phrases
5. Collocations:
   (a) contiguous
   (b) gramrels
6. Sentences

461	win	1	VERB	Colls C	Colls GR	GPs	Bundles	Phrases	Sentences
462	window	1	NOUN	Colls C	Colls GR	GPs	Bundles	Phrases	Sentences
463	windy	1	ADJEC	Colls C	Colls GR	GPs	Bundles	Phrases	Sentences
464	wish	1	VERB	Colls C	Colls GR	GPs	Bundles	Phrases	Sentences
465	with	1	PREPO	Colls C	Colls GR	GPs	Bundles	Phrases	Sentences
466	wonderful	1	ADJEC	Colls C	Colls GR	GPs	Bundles	Phrases	Sentences
467	word	1	NOUN	Colls C	Colls GR	GPs	Bundles	Phrases	Sentences
468	work	1	NOUN	Colls C	Colls GR	GPs	Bundles	Phrases	Sentences

TransformELT

VOCABULARY LISTS
in the CNEC

言

Through the eyes of
CONYE

Let’s look at the database

A work-in-progress

Currently offline but will be available from late February 2023.

Missing words (lempos)

1	back	1	noun
2	back	1	verb
3	back	1	adj
4	back	1	adv

	CNEC	CNEC with conversion	CONYE Corpus
Primary (lempos)	**778**	**1,797**	**4,555**
• Nouns	350		2,791
• Verbs	149		739
• Adjectives	108		553
Junior Secondary	**2,326**	**3,287**	**26,654**
• Nouns	1,120		16,398
• Verbs	532		4,082
• Adjectives	342		4,508

Word families

KS1	smartie	n	256	
KS1	smart	j	27	1
KS2	smart	j	368	1
KS2	smartie	n	131	
KS2	smartly	a	46	
KS2	smarter	a	30	
KS2	smart	v	25	1
KS2	smartphone	n	17	
KS2	smart	n	12	1
KS2	smarttruck	n	10	
KS2	smarten	v	7	

KS = Keystage
KS0 ⇔ pre-primary
KS1 ⇔ primary
KS1 ⇔ primary
KS2 ⇔ lower secondary

1	back	1	noun
2	back	1	verb
3	back	1	adj
4	back	1	adv

KS1	careful	j	242	1
KS1	care	v	30	1
KS1	carefully	a	28	
KS2	careless	j	52	1
KS2	caretaker	n	41	
KS2	carefree	j	10	
KS2	carer	n	5	

nouns modified by "care"	
home	41 •••
facility	11 •••
review	5 •••
package	3 •••
shift	3 •••
worker	3 •••
service	3 •••
work	3 •••
erm	3 •••
cos	3 •••

Native speaker pre-school children

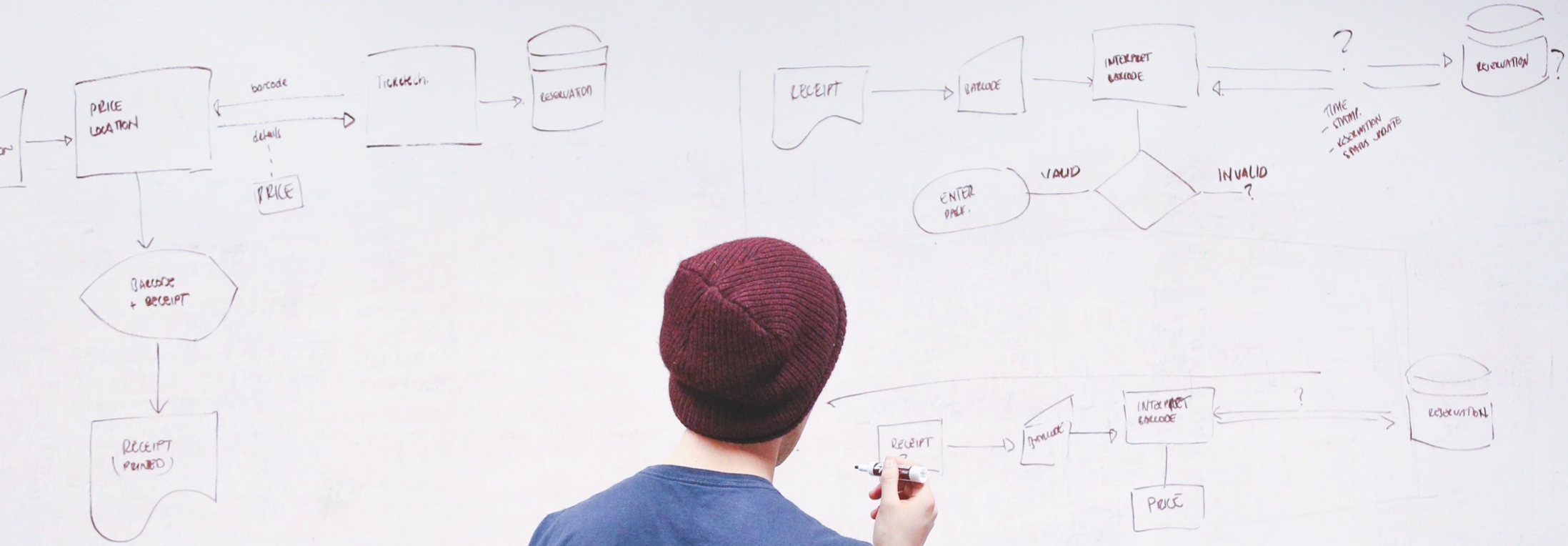
- Estimates of children's vocabulary size vary widely
- In CONYE: up to 5 years old, 11,095 lemmas.
- 2,464 lemmas are in CNEC lists
- Out of the most frequent 500 words that are not in the CNEC list, I chose these 80.

darling, naughty, engine, brick, cream, jigsaw, tractor, sausage, puzzle, toast, shopping, dinosaur, trouser, crocodile, giraffe, spider, poorly, button, panda, crash, track, bite, nursery, frog, tunnel, pretend, monster, bubble, clap, pretend, ladder, helicopter, aeroplane, noisy, brilliant, ambulance, slipper, cuddle, garage, telephone, cupboard, upstairs, stuff, jam, sticker, sort, usually, castle, hiding, jumper, crash, bump, stair, painting, favorite, washing, lucky, roof, blanket, bucket, lamb, clown, swing, drawing, horrible, messy, sleeping, raspberry, pajamas, bedtime, carpet, puppy, normally, scary, chimney, thumb, trunk, yuck, kangaroo, pillow.

Missing words from Junior secondary

- 31,209 lemmas in JS corpus
- Out of the most frequent 500 words that are not in the CNEC list, I chose 140. Words in bold are in pre-school list too.
- **Body:** chin, breathe, skin, bone
- **People**: mum, professor, leader
- **House**: stair, castle, entrance, roof
- **-ly adverbs**: carefully, quickly, immediately, suddenly
- **General**: stuff, somehow, beyond, okay, dot, contain
- **Kids words**: sword, captain, adventure, creature, arrow, blanket, painting, evil, thief, moody, scary, creep

dad, mum, stare, professor, suddenly, whisper, gonna, toward, quickly, finally, grab, slowly, edge, sort, captain, realize, chest, skin, **stuff**, creature, shadow, tear, sight, slip, crowd, guy, silence, thanks, breathe, disappear, cheek, quietly, memory, longer, sword, stupid, **stair**, direction, chapter, ghost, witch, closer, **castle**, slightly, completely, bone, clearly, somehow, weird, roof, okay, carefully, beyond, spot, escape, sharp, distance, definitely, immediately, certainly, apart, indeed, whose, entrance, expression, horrible, **track**, creep, dare, approach, obviously, tight, lucky, eventually, ceiling, fist, arrow, spot, strike, giant, tongue, eyebrow, **bite**, fault, lower, forehead, recognize, feather, bar, bother, type, system, chin, **blanket**, adventure, movement, powerful, properly, **button**, perfectly, **engine**, strength, anymore, beard, monster, whisper, **brilliant**, excitement, curtain, softly, laughter, **painting**, easily, extra, possibly, **sort**, absolutely, remove, nowhere, church, evil, earlier, exist, further, thief, chase, mostly, interested, moody, contain, dot, assume, **spider**, march, confuse, **scary**, leader, interrupt, image, release



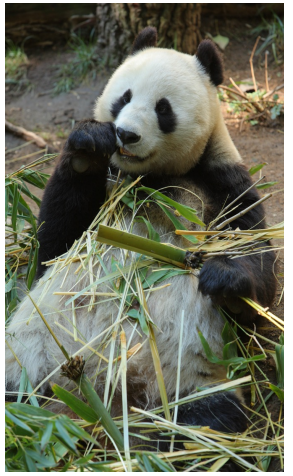
What we do with this data depends on our …

view of language and how we combine words and phrases to create spoken and written discourse.

Not quite a text about pandas

A string of messages

Giant pandas face serious problems in the wild.
It is very difficult for pandas to have babies,
Many baby pandas die when they are very young.
Giant pandas live mainly on a special kind of bamboo.
The bamboo forests are becoming smaller and smaller.
Pandas may not have a place to live or food to eat.



Not quite a text about pandas

A string of messages

As a result

and

Sadly

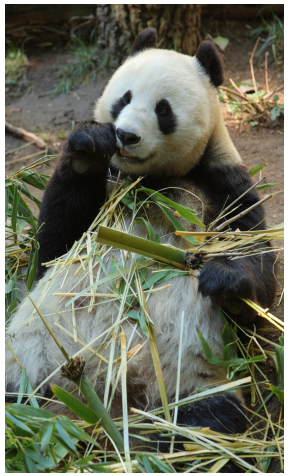
However

For example

also

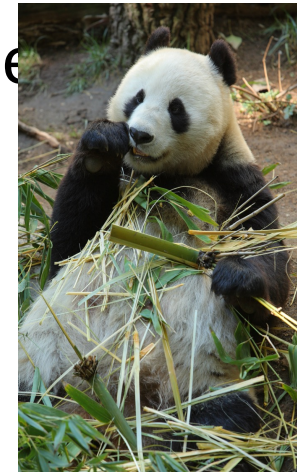
Gramrels

Giant pandas face serious problems in the wild.
It is very difficult for pandas to have babies,
Many baby pandas die when they are very young.
Giant pandas live mainly on a special kind of bamboo.
The bamboo forests are becoming smaller and smaller.
Pandas may not have a place to live or food to eat.



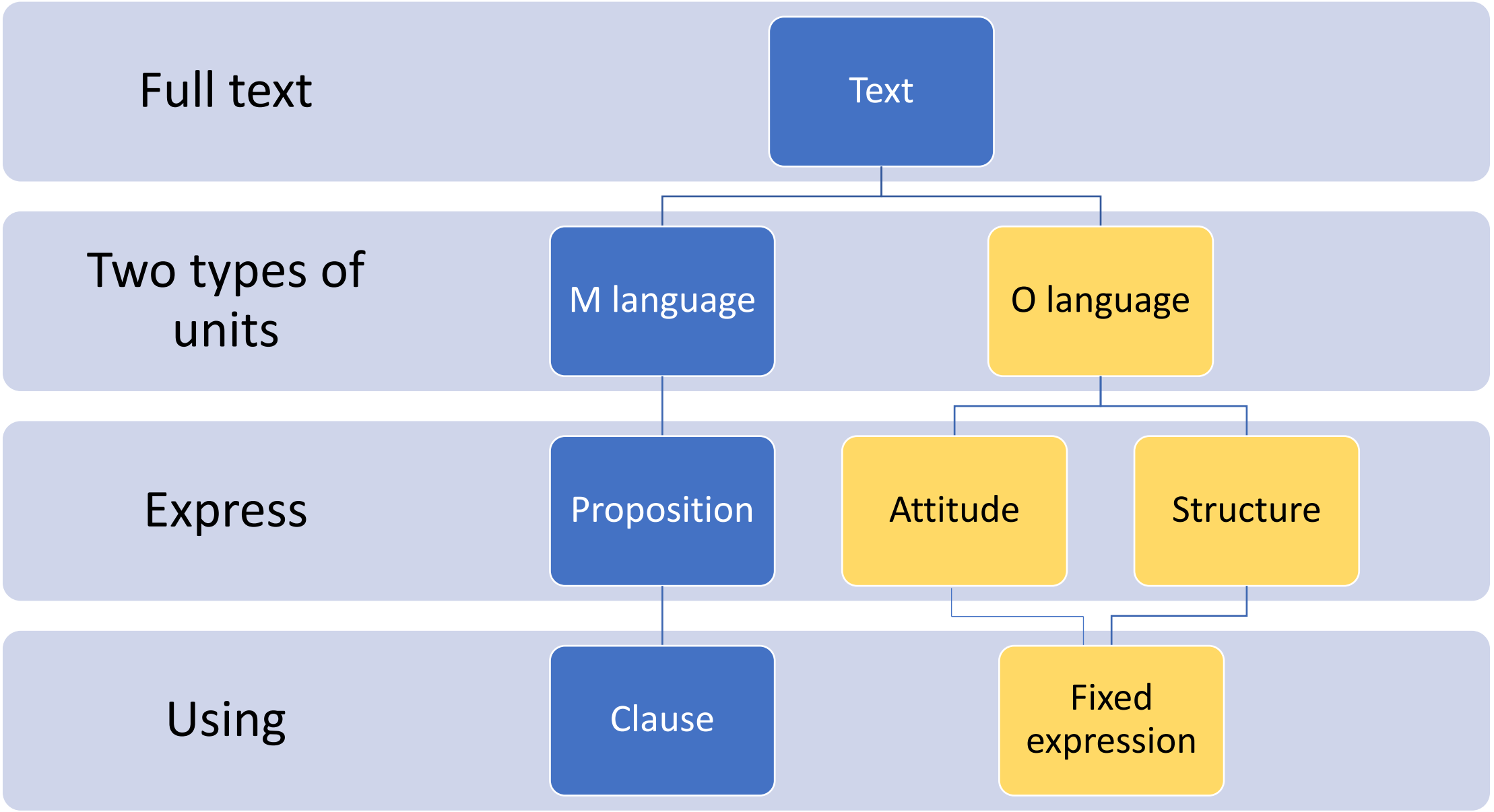
Now it’s a text about pandas

Sadly	giant pandas face serious problems in the wild.
For example	it is very difficult for pandas to have babies,
and	many baby pandas die when they are very young.
Also	giant pandas live mainly on a special kind of bamboo.
However	the bamboo forests are becoming smaller and smalle
As a result	pandas may not have a place to live or food to eat.



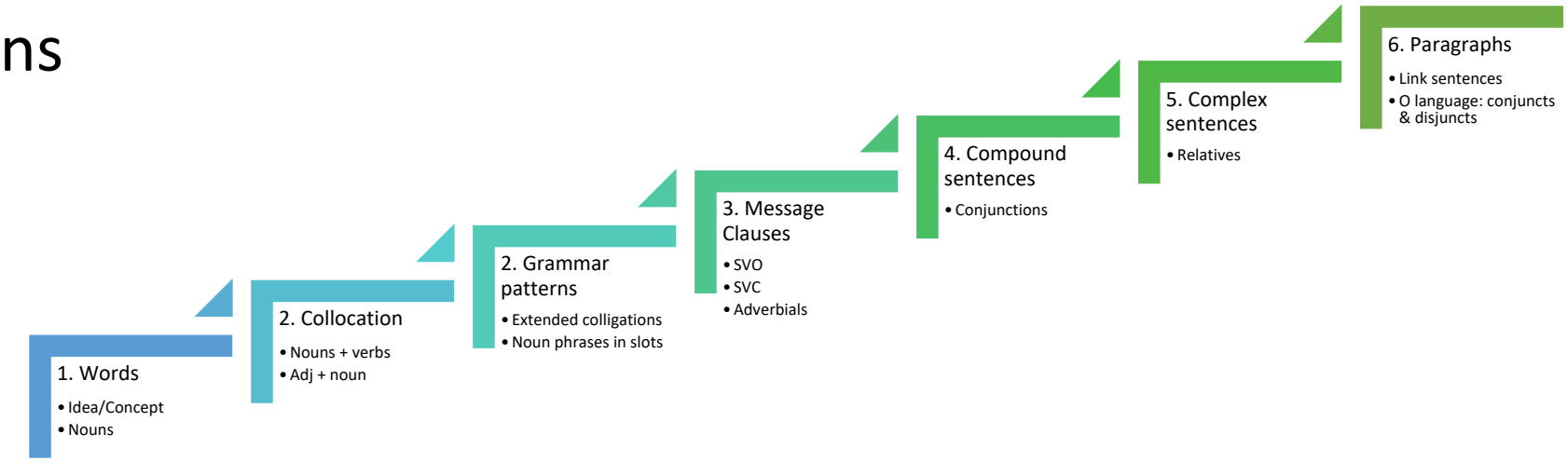
About text

Organisation & Orientation language	Message
Sadly,	giant pandas face serious problems in the wild.
For example,	it is very difficult for pandas to have babies
and	many baby pandas die when they are very young.
Also,	giant pandas live mainly on a special kind of bamboo.
However,	the bamboo forests are becoming smaller and smaller.
As a result,	pandas may not have a place to live or food to eat.


Full text
Text
Two types of units
M language
O language
Express
Proposition
Attitude
Structure
Using
Clause
Fixed expression

A bottom-up, psycholinguistics process

1. CNEC words in all POS
2. Grammar patterns
3. Bundles
4. Phrases
5. Collocations:
   (a) contiguous
   (b) gramrels
6. Sentences



How people use words

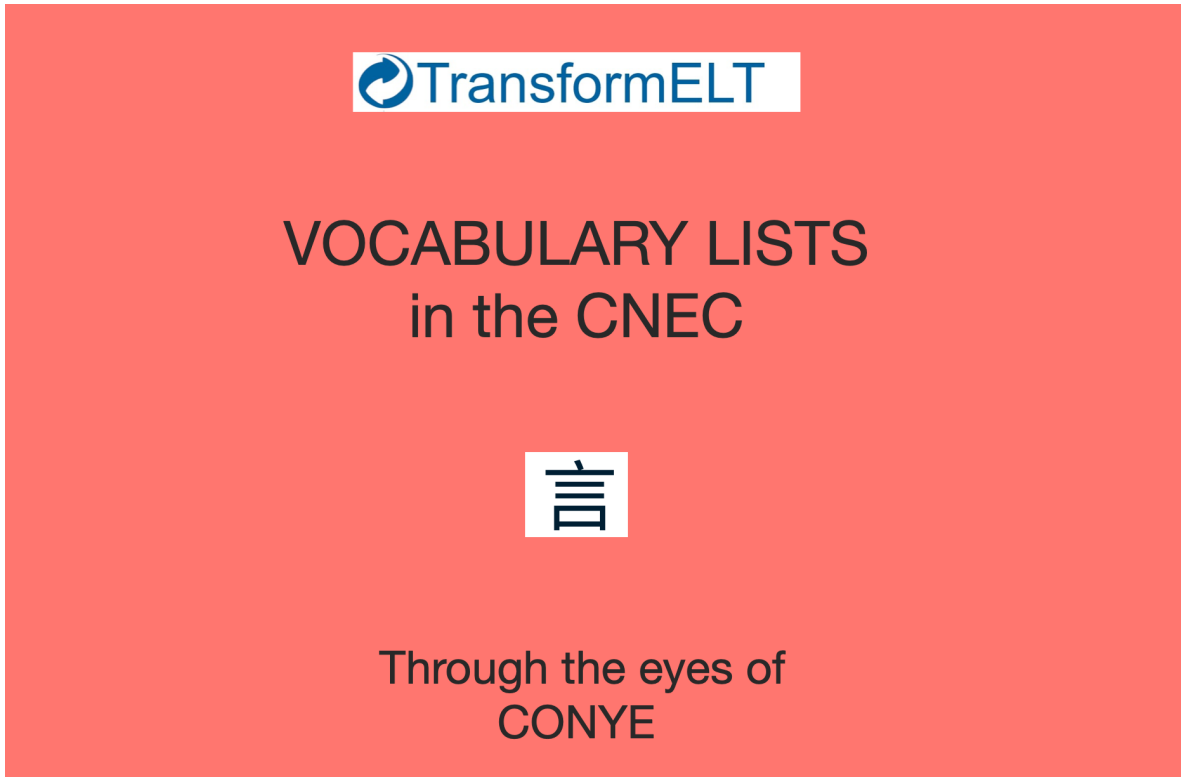
Native-speakers

- Subconscious in fluent speech and writing
- Conscious choices when speaking and writing carefully

Non-native speakers

- Conscious choices of words and combinations: collocation and colligation.
- The more advanced, the more subconscious
- The results of exposure / frequency /patterning.
- Patterns of words / sentences / texts.

The finished product



- A book of lists
- A searchable online database

TransformELT

EMaDA & RIA Interim Progress Sharing Event

James Thomas

versatile.pub@gmail.com

Alan Pulverness

Sarah Mount

10/12 January 2023

Supported by

BRITISH COUNCIL

TransformELT

EMaDA & RIA Interim Progress Sharing Event

Alan Pulverness

Sarah Mount

Alan Mackenzie

John Knagg

https://transformelt.com/

Supported by

BRITISH COUNCIL

End