Producing and comprehending language in monologue and dialogue

Martin Pickering
University of Edinburgh
DOCTOR
The spy saw the cop with the revolver yesterday.
Psychology of language focuses on monologue (and unusual monologue at that)

- Theories of decoding and encoding stimuli in isolation
Why monologue?

- Compatible with traditional cognitive psychology (e.g., visual perception)
  - Psycholinguistics as subfield of cognitive psychology

- Compatible with mainstream theoretical linguistics (e.g., Chomsky)
  - Written (literary) bias
  - Grammaticality judgements on isolated sentences

- Easy to control experimentally
  - e.g., easily measure time to respond
A: I shook hands with Rubenstein once? [and his hand

B: [Yeah we did together

A: That’s right. we were together. wasn’t it incredible?

B: (laughing) oh it was like a cushion.

C: What’s this?

A: [I (0.5)    we shook] hands with Rubenstein.

B: [Rubenstein’s hands.  ]

D: and he had --?

A: his hands –

D: short stubby hands?

A: they were like (0.5) [jelly. they were like – (1.0)

B: [a famous concert pianist

A: they were like (0.5) putty. (0.5)

D: [really?

A: [just completely soft and [limp

B: [mush

A: just mush. it was as though there was [no bone

B: [and warm.
Why study dialogue

Dialogue is the “basic” site of language use

– Antedates monologue in childhood and evolution

– And doesn’t require special training, unlike giving speeches (or even listening to them)
But is it too “messy” to study experimentally?

– How can you control timing when more than one person is involved?

– How can you stop people saying what they like?

Do we need to leave it to discourse linguists, anthropologists …?

I argue, No
How to study dialogue

- It involves many (well-coordinated) cognitive processes
  - Lexical access, parsing, ability to deal with concurrent input, task-switching, turn-taking
  - Many of which are also needed in monologue

- But it is also affected by social factors
  - e.g., influences of interlocutors

- So it should be amenable to psychological science (cognitive + social)
My goal is to understand language processing across the “dialogic continuum”
- From monologue to highly interactive dialogue
- Some processes may not vary across the continuum
- But others may

Dialogue involves concurrent production and comprehension
- Overlapping speech, or 0ms between turns

So production and comprehension are closely integrated, unlike “textbook” accounts

I now explain processing in dialogue, then show how it relates to monologue
Why is dialogue so easy?

“Should” be harder than monologue
- Dealing with changes on the fly
- Can’t always plan ahead
- Working out precisely when to speak and who to speak to
- Producing and comprehending at same time (feedback, overlapping speech)
- Constant task-switching

Two parts to my answer:
- (1) Alignment in dialogue
- (2) Prediction in dialogue

Garrod & Pickering (2004), TiCS
(1) Alignment in dialogue
Interactive-alignment model

Natural dialogue is “messy”, yet interlocutors succeed in reaching understanding

They repeat each other’s words, syntax (grammar) and – ultimately – align their situation models
– Representations of space, time, causality, characters, intentions

Alignment occurs without explicit negotiation
– No evidence of continuous reasoning about each other’s mental states

Rather, alignment “percolates” through levels of representation
– alignment at one level of structure enhances alignment at other levels of structure
– e.g., lexical alignment → syntactic alignment

Pickering & Garrod (2004), BBS
Traditional account

A

Situation Model

Semantic representation

Syntactic representation

Lexical representation

Phonological representation

Phonetic representation

B

Situation Model

Semantic representation

Syntactic representation

Lexical representation

Phonological representation

Phonetic representation
Interactive-alignment model

A

Situation Model

Semantic representation

Syntactic representation

Lexical representation

Phonological representation

Phonetic representation

B

Situation Model

Semantic representation

Syntactic representation

Lexical representation

Phonological representation

Phonetic representation
Alignment at different levels

- Referring expressions (Brennan & Clark, 1996)
  - A and B are discussing a meal at a restaurant
    - A says “chef” → B also says “chef”
    - A says “cook” → B also says “cook”

- Situation models (Garrod & Anderson, 1987)
  - A and B are navigating around a maze
    - A: two along two up → B: four along three up
    - A: I’m at B2 → B: I’m at D3

- Alignment on sounds (e.g., Pardo, 2006)
Alignment of syntax

Conversation between bank robbers:

A: *Cor, the noise downstairs, you’ve got to hear and witness it to realise how bad it is.*

B: *You have got to experience exactly the same position as me, mate, to understand how I feel.*

Schenkein, 1980
Alignment of syntax

Conversation between bank robbers:

A: *Cor, the noise downstairs, you’ve got to hear and witness it to realise how bad it is.*

B: *You have got to experience exactly the same position as me, mate, to understand how I feel.*

Schenkein, 1980
Alignment of syntax

Branigan et al. (2000), Cognition
Alignment of syntax

The cowboy handing the banana to the burglar (PO sentence)
The cowboy handing the burglar the banana (DO sentence)

Branigan et al. (2000), Cognition
Procedure

- Confederate describes card: *The chef giving the jug to the swimmer*
- Participant finds the card that matches this description
- Participant picks up next card

- Participant describes card: *“The cowboy handing ...”*
Same vs. different verb

- 4 prime conditions:
  - PO-same: The chef handing the jug to the swimmer
  - DO-same: The chef handing the swimmer the jug
  - PO-different: The chef giving the jug to the swimmer
  - DO-different: The chef giving the swimmer the jug

- Target:
Results: % PO target

![Bar chart showing percentage of PO targets for Same verb and Different verb with PO prime in green and DO prime in gray. The chart indicates a higher percentage for Same verb compared to Different verb.]
- Priming when verb is not repeated

- But stronger priming when it is repeated
  - Just as in monologue (Bock, 1986; Pickering & Branigan, 1998)
  - The **lexical boost** to priming
  - Hence lexical alignment enhances syntactic alignment

- Semantic alignment enhances syntactic alignment

  *the goat that’s pink* → *the sheep that’s pink*

  MORE THAN

  *the book that’s pink* → *the sheep that’s pink*
Thus, alignment at one level enhances alignment at another level.

Such interactions between levels underlie mutual understanding.
Representing this account

Following Levelt et al. (1999)
Give

Conceptual stratum
(meaning)

Lemma stratum
(syntax)

Form stratum
(sound)

GIVE(X,Y,Z)
GIVE(X,Y,Z) 

give 

HAND(X,Y,Z) 

hand 

PO 

DO 

Conceptual stratum 

Lemma stratum
The chef giving the jug to the swimmer
The chef giving the jug to the swimmer

Conceptual stratum

Lemma stratum
The chef giving the jug to the swimmer

give

GIVE(X,Y,Z)

hand

HAND(X,Y,Z)

PO

DO

Different verb (hand)

Conceptual stratum

Lemma stratum

Syntactic alignment
The chef giving the jug to the swimmer

Same verb (give)

Syntactic alignment + lexical boost
Cross-linguistic alignment

Confederate described a picture in Spanish
  – Active: *El taxi persigue el camión*
    - “The taxi chases the truck”
  – Passive: *El camión es perseguido por el taxi*
    - “The truck is chased by the taxi”

Participant made yes/no judgement

Then described a transitive picture in English
Spanish → English priming

Participants produced more English passives after Spanish passives than Spanish actives
Alignment

Many studies show strong syntactic alignment
  - For different constructions, languages, etc.
  - Children, non-native speakers, even aphasics

Corpus studies reveal similar effects “in the wild”

Alignment also occurs in comprehension
  - e.g., interpreting ambiguous utterances

And such alignment is not limited to syntax
Alignment

Priming *within* a speaker in monologue is qualitatively similar to alignment *between* speakers in dialogue

- So these representations and processes *do not differ* between monologue and dialogue
- Priming *may* be stronger in dialogue, as would befit a mechanism of alignment
(2) Prediction in dialogue

- So what mechanisms underlie alignment?

- How do those mechanisms relate to production and comprehension?

- So far we have assumed *parity*
  - Production and comprehension both activate the same representations (e.g., lemma nodes)

- But how can we explain other aspects of dialogue such as its fluidity?
A: I shook hands with Rubenstein once? [and his hand
B: [Yeah we did together
A: That’s right. we were together. wasn’t it incredible?
B: (laughing) oh it was like a cushion.
C: What’s this?
A: [I (0.5) we shook] hands with Rubenstein.
B: [Rubenstein’s hands.  ]
D: and he had --?
A: his hands –
D: short stubby hands?
A: they were like (0.5) [jelly. they were like – (1.0)
B: [a famous concert pianist
A: they were like (0.5) putty. (0.5)
D: [really?
A: [just completely soft and [limp
B: [mush
A: just mush. it was as though there was [no bone
B: [and warm.
Fluidity

- Usually very little interval between turns (200-500ms) (Stivers et al., 2009)
  - Even though it takes 600+ ms to plan simple utterances (e.g., Indefrey & Levelt, 2004)

- Interlocutors complete each other’s utterances, often without delay
  - And have similar (aligned) understanding
    - my completion expresses roughly the same meaning as yours would (see limp vs. mush)

- We account for this fluidity through prediction
An integrated model

- **prediction**: determining what you or your interlocutor is likely to say next

Production and comprehension are tightly interwoven

- Prediction during production involves comparing predicted and actual self-comprehension (“listening to yourself”)
  
- Prediction during comprehension involves simulating aspects of the way you would produce the utterance

Pickering & Garrod (2013), *BBS*
Predicting your own actions

When I plan an arm movement to a target (e.g., for grasping)

- I formulate an action (motor) command to move my arm
- I move my arm
- I construct a percept of the experience of arm movement
Predicting your own actions

But also, when I plan this arm movement

- I construct an efference copy of the production command
- This inputs into forward models which generate the predicted act of arm movement
- And the predicted percept of the experience of arm movement

These predictions are (generally) ready before the actual arm movement and its percept (e.g., Wolpert, 1997)

- They may use equations (e.g., hand coordinates) and approximations
- The actual movement and the prediction can be compared
Predicting your own speech

- We extend this account to speech

- When I speak, I formulate a production command, speak, and experience the act of speaking

- But I also construct an efference copy of the production command that is used to predict the act of speaking and the associated experience
  - I separately predict semantics, syntax, and sound ("phonology")
  - I can compare these predictions with my actual representations used in production
  - This is known as "self-monitoring"
production command → implementer

efference copy → utterance percept

predicted utterance percept

forward models → monitor
Self-monitoring

I construct my predictions of what I will say (at different linguistic levels)

Then I prepare my utterance (via implementer)

And I can then compare my predicted percepts with my actual percepts

- Predicted semantics with semantics
- Then predicted syntax with syntax
- Then predicted phonology with phonology

So-called “production-based” self-monitoring
Example

To the ye- to the orange node (Levelt, 1983)

- Speaker computes predicted utterance percept /ɒ/ (via forward model)
- Based presumably on history of uttering /ɒ/ after planning to describe
- Speaker computes utterance percept /j/ (via implementer)
- Monitor compares /ɒ/ with /j/ and determines discrepancy
- And speaker reformulates (on this occasion)
Predicting other people’s actions

I can predict other people’s actions based on my experience of other people’s actions

- *prediction-by-association*
  
  - e.g., how they have moved their arms in the past
  
  - like predictions of inanimate objects (e.g., falling rocks)

But I can also predict other people’s actions based on my experience of *my own* actions

- *prediction-by-simulation*
  
  - e.g., how I have moved my arm in the past
Roughly, “what would I do under the circumstances?”

- I see start of arm movement
- And then determine how I would move “if it were my arm”
- In other words, I *covertly imitate* the actor

I use this covert imitation as the basis for action prediction

Similarly, I hear the start of the speaker’s utterance

- I covertly imitate the speaker and use this as the basis for linguistic prediction
- I can then compare my prediction with my representations of the speaker’s utterance when it is produced (“other monitoring”)
utterance (time $t$) \rightarrow \text{utterance (time $t+1$)}
Example

You utter *I want to go out and fly the*

I covertly imitate your utterance so far to derive:

- A representation of your utterance so far
  - i.e., the meaning, syntax, and sound of *I want to go out and fly the*

- the production command that I would use to produce this utterance

- the production command that I would use to produce the predicted next word “kite”

I use an efference copy of this second production command to predict aspects of the meaning, syntax, and sound of “kite” (i.e., KITE,noun,/kalt/)
I may act on these predictions
- e.g., look for a kite (cf. Altmann & Kamide, 1999)

I can compare my predictions with what you then say
- Again, the predictions are normally ready before the utterance occurs
- This is “other monitoring”
- Such monitoring constantly occurs during comprehension
- If the predictions are wrong, I may resort to general (non-linguistic) reasoning

Pickering & Garrod, (2014), Frontiers in Human Neuroscience
Alignment and prediction

Alignment supports prediction:
- predictions are better when we are well-aligned

And covert imitation underlies both prediction and alignment
I have shared representations of my own and my partner’s production mechanisms.

The mechanisms should work appropriately when used sequentially:

- I use them when I’m talking
- I use them again when you’re talking

But they may be overburdened when used concurrently.

To control for irrelevant feedback, we used *imagined* production.

Gambi et al. (2015), *JEP:LMC*
Participants John and Mary are in adjacent sound-proofed rooms (with a viewing window).

They name pairs of pictures, starting with the one in the designated colour.

They are also told their partner’s task:
- Name pictures in the same order
- Name the pictures in the other order
- Don’t name the pictures

We measured time to start naming.
John: BLUE

Mary: BLUE
John: BLUE

Mary: RED
John: BLUE

Mary: NO
Participants take longer to initiate naming when they believe their partner is also naming (around 10-15ms)

- Though not when they believe their partner is categorizing (same vs. different category)

Participants unaffected by what they believe their partner is saying (i.e., the order of naming)

They seem to be predicting their partner’s act of naming

Prediction is used “off-line” to imagine language (Gambi & Pickering, in press, LCN)
Do speakers use production mechanisms to predict others’ utterances?

Participant names Picture 1, but occasionally (12%) it changes into Picture 2

She stops Picture 1 later when she then has to name Picture 2 than when she only has to stop

Mary: *ap* –

Mary: *apple chair*

Hartsuiker et al. (2008), *Cognition*  
cf. Tydgat et al. (2011), *JML*
Do speakers use production mechanisms to predict others’ utterances?

A participant also stops Picture 1 later when *her partner* then has to name Picture 2 than when she only has to stop.

Mary: *ap* - John: [nothing]

Mary: *apple* John: *chair*

Gambi et al. (in press, *Cortex*)
Speaker appears to predict partner’s upcoming act of production

And this interferes with speaker’s own act (stopping speech)

- Representation of “speak” code interferes with “stop” code, even if the “speak” code is for another person

- An effect of prediction primarily relating to timing

A single task (here, stopping-restarting) is distributed across two people

Behaviour in joint task is similar to behaviour in solo task
These studies suggest that speakers represent their partners during turn-taking:

- And that they use these representations to make predictions
- In a way that facilitates the fluidity of dialogue
Turn-taking in dialogue

- Comprehender predicts speaker by simulating utterance and running on ahead
  - Using mechanisms associated with production
  - And can therefore concurrently construct response

- This is effective when comprehender is well-aligned with speaker
  - Because comprehender’s predictions will tend to be accurate

- These predictions are combined with entrainment on speech rate to aid fluidity

Garrod & Pickering, in press, *Frontiers*
Conclusions

Dialogue is facilitated by alignment and prediction

But many mechanisms are the same as the mechanisms used in monologue

- priming in monologue ⇔ alignment in dialogue
- predicting your own utterances ⇔ predicting your partner’s utterances
Conclusions

- Psychology of language needs to understand the whole “dialogic continuum”

- An integrated psychology of language should be grounded in the study of dialogue

- But it can and should also be applied to monologue
Acknowledgements

Sarah Bernolet
Holly Branigan
Zhenguang Cai
Ciara Catchpole
Sandie Cleland
Uschi Cop
Albert Costa
Nivja de Jong
Vic Ferreira
Chiara Gambi

Simon Garrod
Rob Hartsuiker
Janet McLean
Claudine Raffray
Sofie Schoonbaert
Michael Stevens
Ilse Tydgat
Joris van der Cavey
Eline Veltkamp
Hao Yan