Psycholinguistic and neurolinguistic evidence for a rich model of linguistic representation
or, why linguistic representations must be redundant

Abstract
Theories of linguistic representation have been shaped – and handicapped – by a methodological imperative and two invisible and unexamined assumptions.

• The imperative is:
  o be parsimonious
  o minimize redundancy.

• The assumptions are:
  o The representation of a word in the ‘mental lexicon’ is a structured list (like an entry in a printed dictionary); this ‘lexical entry’ does not include the wider range of information that would be included in an encyclopedia.
  o The mental representation of the structure of a complex word or a sentence is like a two-dimensional diagram on a page or blackboard: it is either ‘present’ or ‘absent’, rather than having gradient strength.

• However, gradience is implied by a usage-based view (exemplar model) of the individual’s language knowledge, because in a usage-based model, more experience with a word or a syntactic structure gives it a stronger representation. In particular, more frequent words and structures should have stronger representations (other factors being equal).

• A usage-based model also implies redundant representation: words and structures are usually learned in rich settings, and work in psychology indicates that vastly more ‘setting’ information remains stored in our brains than parsimony-driven theories predict.

• Data from the aphasias and speech errors further support the idea that linguistic representations should be thought of as multi-dimensional, gradient, redundant, and linked to non-linguistic experience. Error and preservation patterns in aphasic speech show that the brain makes use of the frequencies of words, constructions, and collocations, as well as category membership and hierarchical structure, during language processing.

• Frequency effects are evident along two quasi-independent axes: syntagmatic (the sequential context, e.g., deploying correct functors, categories, and utterance-level intonation) and paradigmatic (the choice at any given linguistic level, e.g., selecting content words and modifying structures).

• Language production and comprehension require constantly settling competition among possible outputs/interpretations; the difficulty of settling on a semantically, syntactically, and pragmatically correct form is affected by sequential and simultaneous co-occurrence probabilities across linguistic levels. These claims are supported by the literature on experimental work in normal speakers.

• An account of language representation and processing that encompasses frequency as well as categorization and structure is compatible with what we know about how the brain works: increased experience with a linguistic structure results in increased activation—and strengthening—of the neural networks involved in processing that structure. The substantial redundancy in representations and processing that is introduced by incorporating both frequency-based and hierarchy-based information is in fact appropriate for the brain as a fast, reliable, massively parallel error-correcting network with very large storage capacity and gradient representation strength. A language processing model that includes these types of information has the potential to resolve some of the tensions between usage-based and ‘structure-based’ theories of mental representation – in particular, how frequency of occurrence and of co-occurrence of words, morphemes, structures, and constructions can be an integral part of our mental representation of language.
Prologue: This is not a pipe. It’s only Magritte’s picture of a pipe.

- Formalisms/notations (‘S > NP + VP’)/any way of describing a language), similarly, are not mental grammars. They are not what we have in our brains, in other words. They are only maps, metaphors, external representations of our mental grammars.
- Notations/formalisms/descriptions are, nevertheless, important because they constrain the way we think about the world (including the hidden world inside our heads).
- We linguists (functionalists as well as formalists) have some bad habits of thought caused by the notations and diagrams that we use.
- I present a (mini)-model for language production that affords less-constraining, neurally-plausible ways for linguists and cognitive scientists to think about the mental representation of language.
- It’s NOT a computational model – that would be more complex.
- (And it’s not the real thing; it’s only a map.)

What’s wrong with these two standard assumptions?

1. The representation of a word in the ‘mental lexicon’ is a structured list (like an entry in a printed dictionary); this ‘lexical entry’ does not include the wider range of information that would be included in an encyclopedia.
2. The mental representation of the structure of a complex word or a sentence is like a two-dimensional diagram on a page or blackboard: it is either ‘present’ or ‘absent’, rather than having gradient strength.

For starters, they have no plausible relationship to what we know about how information gets into our brains, and to how representations of that information are formed.

Our minds store information in structured local and long-distance networks that are built up and pruned down by experience, starting from a pre-wired anatomical base. We can learn about what those networks are like from experiments, analyzing error corpora, and analyzing language in learners and in people with aphasia. So linguists need to know the basics of psycholinguistics and neurolinguistics.

“It makes no sense to say that language is in our minds and then to deny a priori that principles which apply everywhere else in the mind do not apply to language (Menn & Duffield 2013).”

A few more things that functional linguists need to know about how our brains work

- Information spreads in parallel across these networks, and networks interact.
- Processing is highly redundant, and encoded information is very rich – parsimony has no role in the design of brain networks (except that connections that are not used become weak and effectively disappear). For example, we subconsciously remember things about the settings in which we heard particular utterances.
- Experience basically works on Hebb’s principle of association formation: “What fires together, wires together.”
- Activation, spreading automatically in our brains, produces major but unconscious influences on our choices of referring expressions, word order, and syntactic structure.
Spreading activation

More things that linguists need to know about how our brains work: Spreading activation, priming, interaction

- Activation spreads automatically from one neuron to others that handle similar or sequentially related information.
- Priming is a rise in the level of activation of a particular neural structure due to the automatic spreading of activation from some other neural structure:
  - Lexical priming is the spread of activation to particular words when related words have been aroused, or the maintenance of activation on a word once it has been aroused.
  - Structural priming is the maintenance of activation of particular syntactic structures that have recently been used; syntactic structures might also prime structures that are similar, not just ones that are identical.
- When we talk, activation comes from many sources, including what we intend to communicate and what we have heard. Interaction (at multiple production 'stages') between these sources creates the eventual utterance.

Quotes from Patricia Clancy, “Dialogic Priming and the Acquisition of Argument Marking in Korean” (2009)

- “...even if re-use of a form is functionally motivated and serves semantic, pragmatic, or interactional functions in the discourse itself, involuntary priming can also have a facilitating effect on production. Thus priming and discourse–functional motivation can be seen as complimentary forces that work together to create resonance.” (pp. 106–7)
- “...I will assume that ... the grammatical morphemes in an utterance, such as those indicating important syntactic functions and grammatical relations, can also act as structural primes for subsequent use of the same morphemes.” (p. 107)

Part 1: Data from the aphasias and speech errors

How language disorders/errors can tell us something about the linguistic knowledge represented in our brains: the 'black box' (empirical/bottom–up) approach

- Mistakes indicate difficulty in language processing.
- Analyzing these errors provides information about what the source of the speaker’s difficulty might have been.

Then we can ask:

- What sort of linguistic/cognitive processing system is it that would find these particular tasks difficult and other tasks easy?
- and, in such a system, How is linguistic knowledge represented?
Constructions & collocations, aphasia & experiments

Constructions and collocations are both sequential structures

- **Construction**: a form–meaning pairing (regardless of frequency),
  *e.g.* single morphemes, inflected forms, compounds, phrases, and clauses.
- **Collocation**: a sequence of words that co–occur frequently,
  *e.g.*, *It’s a*, *gimme*, *suppose we*, *ladies and gentlemen*, *wash the dishes*,
  regardless of whether that sequence forms a syntactic unit or has a well–defined meaning.

Pourquoi l’aphasique peut–il dire: “Je ne peux pas le dire” et pas “Elle ne peut pas la chanter”? (Nespoulous & Lecours 1989) [Why can the aphasic person say I can’t say it but not She can’t sing it ?]

- Possible culprits:
  - lexical frequency
  - collocation frequency,
  - *formula* (= construction with lexical specification) *I can’t VERB it*
  - emotional weight

- Dressler’s (1991) work on Breton: A speaker with fluent aphasia tends to name pictures or examples of a single object using:
  - the **plural** form if the object itself is most frequently found in quantity (*leaves*, *potatoes*)
  - the **dual** if the object is usually found in pairs (*eyes*, *hands*).

- Similar preference for plural when asked to name a singular picture occasionally found for *eyes*, *hands* in aphasic English speakers.

- Experimental work (e.g. Baayen & colleagues): Normal speakers are relatively faster at processing the plural when a word is usually used in the plural form.
  - So even transparent ‘composite’ forms can be easier to process, if they are relatively frequent as compared to their stems.

Sequential structure in language production: evidence from aphasia

- Utterance–initial collocations in almost all aphasic speakers
  *There’s a..*  
  *It’s…*  
  *It’s a…*

- Long runs of low–content collocations in severely–impaired fluent aphasic speakers
  - *...they’d’ve been put to [nonwords] and made [nonword]…;
  - *I think that there’s an awful lotta* [nonword], but *I think I’ve lotta net and tongued it*.

- Shorter collocations in (non–fluent) agrammatism:
  - *... forgot the wash the dishes*
    - Target: ‘forgot that she was washing the dishes’
  - *... I like the – go home.*
    - Target: 'I’d like to go home.'

(Menn, 1990; Buckingham & Kertesz, 1976)
The syntagmatic axis: ‘Agrammatic’ aphasic speakers show the effect of high sequential probabilities in collocations

...forgot the wash the dishes... ‘...forgot (that) she was washing the dishes’
...I like the – go home. ‘...I’d like to go home.’

• After she has chosen the definite article ‘the’ to follow forget or like, this speaker with Broca’s aphasia is in trouble.
• She ‘plugs in’ familiar phrases (wash the dishes, go home) with appropriate semantic content –
• But those phrases cannot grammatically follow the...
• People with moderate Broca’s aphasia – including this speaker – do not generally make such gross errors as substituting the article for the infinitive marker.
• like the and forget the are reasonably common as sequences

But: the collocation ‘V+the’ goes across the major syntactic boundary between the verb and what should be the start of its NP object. How can sequences like this play a role in sentence production? Hold that thought.

The frequency of a construction is important

An example of frequency/structure interaction:

• ‘Shrink’ is more frequently used in the undergoer–subject argument structure The sweater shrank two sizes
  than in any other, e.g., They shrank the sweater two sizes
• Relative verb–(subcategorization) frame frequencies create an expectation that affects readers’/listeners’ processing patterns and comprehension:
• People with aphasia comprehend sentences better when the verb is in its most frequently used frame. (Gahl et al. 2003)
  o This supports a processing model that has relatively direct semantic construal of the verb frames of simple clauses, as construction grammars would predict.

So: If frequency is represented in our minds, how can we capture that fact in a construction–based grammar?

Mental representation needs to include constructions (form–meaning pairings)

• Entrenched constructions comprise a wide spectrum of instances
  o from formulaic expressions and idioms (on the highly–specific end of the spectrum) to open formulas like Subject–Verb–Object > Agent–Action–Undergoer (on its most general end).
• Frequency of use creates entrenchment.
• In other words, constructions that a speaker frequently uses become his or her automatized neural or neuromotor routines
• A speaker’s automatized routines are deployable as single units in processing.
• NB: as you know, claiming that a construction is a unit is not the same as claiming that it is unanalyzed.

Mental representation needs to include collocations

• Collocations also become entrenched with use:
  • {Can I [have] (some] ice cream)?
  • {I (wonder [if) {there’s] (some] ice cream) [(left over) [from]) (the] party) last night).
Marking the constructions and collocations results in improper brackets galore, adds redundancy – although we don’t yet have the tools for defining collocation, which must be a gradient notion. We need to develop them.

**The paradigmatic axis: Our minds use hierarchical and categorial information in language production – even in producing idioms and lexicalized phrases**

- Blend’ speech errors (common in normal speakers as well as speakers with aphasia) show that even highly automatic collocations may have internal structure for users, because these **errors happen at structural ‘joints’, preserving grammatical category** – at the expense of semantics and of how likely the sequence of words is.
- **My welcome.** *(My pleasure/You’re welcome)*  
  *Oh, help all you want. (Help yourself/Take all you want)* (spoken at the dinner table)
- **I lost my track of thought.**
- **The problem is to find off a piece you can chew.**
- **I really liked the original Pink Panda.** *(discussing the movie Kung Fu Panda)*

**Genre expectations add competition to word arousal in narrative**

Folk–tale genre expectations, combined with the sequential expectations of what is likely to come after ‘wicked’, cause lexical competition that this agrammatic aphasic participant controls only with conscious effort. Here, the wicked **wolf** in ‘Little Red Riding Hood’ becomes a wicked **witch**, then a wicked **fairy** *(Kolk et al., 1990).*

- **Midden in het bos pakte de heks, nee, wolf en Roodkapje is opgegeten**  
  - ‘In the middle of the forest the witch, no, wolf caught and Little Red Riding Hood has been eaten up.’
- **de boze fee de boze wolf in het bed, ja.**  
  - ‘the wicked fairy, the wicked wolf in the bed, yes.’

**Concluding Part 1:**

What language disorders/errors tell us about the linguistic knowledge in our brains is

- that hierarchical structures, constructions, and collocations are **all** represented somehow,
- and that frequency and conditional probability – at all levels and crossing levels – are also represented somehow.

**Part 2:** A conceptual sketch of how knowledge of what’s usually called ‘grammar’ and what’s usually called ‘usage’ can be integrated psycholinguistically.

All of this information is knowledge of language; there’s no reason to exclude any of it from our conception of mental grammar, even if (perhaps) not all of it can be handled in a single kind of formalism.
Errors show that competition is endemic in the brain, and is resolved by many kinds of information coming together

- categorial (e.g., ‘this is a prepositional phrase’)
- hierarchical (‘this prepositional phrase modifies this noun’)
- sequential (‘verbs of motion are likely to be followed by expressions referring to the goal or direction of motion’)
- lexical (‘this is a verb of motion, possible meanings are _____, _____, ....’)
- encyclopedic (‘sweaters are likely to shrink’)...

Competition among different kinds of information is normally resolved in ways that let us understand and create novel utterances, e.g.

Sequential, categorial, and hierarchical information can be integrated in a conceptual production model

- **MISCHA**: A conceptual model for linguists and psycholinguists to use in organizing data and designing experiments
- (Still much simpler than a true neurologically realistic model – cf. Nadeau 2012)
- MISCHA allows linguists to think about the representation of hierarchical grammar and sequential usage simultaneously
- **Analogy**: the structure of music involves harmony AND melody (not to mention rhythm, timbre, etc.)
• **Standard production model (LM version): cascaded & modestly parallel**

  **Cascaded**: as soon as any component is ready, it sends activation out to the next level down.

  **Functional Level** – first linguistic level, unordered, pre-syntactic: consists of aroused lemmas, referent lemmas tagged with their roles, other semantic information

  \{GO\_PERF, MOUSE\_DEF\_S\_OLD = PRO\_NEUTER, corner\_LOC…\}

  **Modestly parallel**: verb & other role-governing lemmas set up frames with slots for their arguments

  tagged forms of referring expressions get assigned to slots

  **Positional Level** – second linguistic level, ordered, syntactic structure; word lemmas are now in appropriate grammatical frames:

  (PRO)\_S\_SUBJ\_NP (GO)\_V\_PERF\_BACK\_ADV\_IN (det CORNER)\_NP\_LOC\_PP

  **Phonological Encoding Level** /ɪts ɡən bæk ɪn ɗəˈkʰɪɾɪ/ womain

  **Speech Gesture Level** – heard as [sgəmˈbækʰɪɾɪˈkʰɪɾɪ]

  **What's missing from the standard model? SEQUENTIAL connections!**

  • **Constructions!**  **Formulas!**  **Idioms!**
  • Automatized phrases and collocations, and the resultant contractions
  • (it’s, it’s gone, ‘sgone, inna)!
  • The standard sentence production model fails to use as much as it could of the massively parallel, redundant computing capacity of our brains.
  • If we invoke spreading activation (which we know exists) and use more of that massively parallel capacity, we can have sequential as well as ‘downward' (node to daughter node) activation at all levels of the model.
  • So we can incorporate e.g. the sequential information which allows the phonological reduction of in the to [ɪnə] – without losing the hierarchical information of [IN [det CORNER]\_NP]\_PP
MISCHA is seriously parallel

- Nodes at every level – not only words, but category nodes – activate others, to varying extents, thus capturing speakers’ expectations (stored as connections of varying strengths) about what is likely to come next.
- Constructions and collocations are represented in the grammar by strong connections between frequently co-occurring words and/or nodes.
- Hierarchical (constituent structure) information is represented by connections between syntactic nodes and their daughters.
- What is a constituent is an empirical matter, but we probably don’t need to have bristly trees, because there is no reason to insist that all grammatical information (e.g. case, tense) be forced into 2-dimensional tree-branch formalism.

MISCHA spots the mouse

It’s gone back in the corner!

Initial parallel activations:

- Message [mouse is in (dark) corner of bookcase].
  Mouse has already been mentioned; location is new information
  (Pointing gesture Is also activated by intention to communicate location)
- Basic constructions for expressing location: There it is!
  (NP + copula + Loc)
- Highly probable ways to start an informative utterance: It’s, It’s a, There’s, I’m, I can…

Converging on a particular start (It’s) has these consequences:

- inhibiting competitor words and structures
- increasing activation of next likely items in construction and in hierarchy
- maintaining activation on the construction and hierarchical structure(s) that are compatible with that start.

Converging on the start It’s also does these things:

- feeds back the information as to what part of the S, C and H that the starting word has satisfied (grayed parts of diagrams below)
- increases activation of the likely next items in construction and in hierarchy (bolded items in diagrams)

- maintains activation on the Sequences, Construction(s) and Hierarchical structure(s) that are compatible with that start.
• Working out LOC
  o two elements of information chosen for expression: [back] in the corner
  o choosing [back] inhibits competitor words and structures
  o & feeds back the information as to what part of the S, C and H that it has already satisfied (grayed parts of diagrams below)
  o increases activation of likely next items in construction and in hierarchy (=bolded items in diagrams)

[utterance (S)]
Locative construction (S):
NP cop Loc direction Loc place

[utterance (H)]

• Activation of the locative prepositional phrase in the corner will come from several sources in parallel – work them out!

How does MISHA handle the problem cases we started with?
• Some aphasic people can say
  I can’t say it. but not She can’t sing it.

Formula, frequently utterance–initial, emotionally loaded:
  I can’t [VP]

Frequent verb (also emotionally loaded, if you are aphasic): say
Super–frequent object NP: it

• A Breton speaker with fluent aphasia tends to name pictures or examples of a single object
  o using the plural form
    if the object itself is most frequently found in quantity (leaves, potatoes)
  o using the dual form
    if the object is usually found in pairs (eyes, hands).

Perfect example of
high transition probability > stored composite form.

What does MISCHA capture about representation?
• Sequential probabilities from one word to the next and from one node to the next are learned, therefore they are represented. They seem to be represented as connection strengths – not as anything we’d find in a dictionary, but more like what drives Google search.
• Probabilities are inherently gradient from weak to strong (i.e. from ‘never’=0, thru all intermediate values, up to ‘always’=1), rather than being either absent (0) or present (1).
• So sequential (syntagmatic) links are gradient – and they can be VERY strong, leading to clitic formation and to lexicalized contractions (whatcha, gimme...), just as our intuitions and observations have been telling us all along. (I) (am (go)ing (to V)> [mənə] V

How can we tell whether the MISCHA conceptual model of representation is reasonable?
It has to make sense of data about how real speakers behave. We’ve focused on the first two data types, but they all are relevant:
• ease or difficulty in producing & understanding patterns in disorders (preservations and losses)
• learning patterns (correct & incorrect)
• usage, including variation
• historical change
• judgements and other meta-linguistic behavior
• normal errors/self-corrections
References and suggested readings for Menn talk, 15 November 2023


