1 Introduction

1.1 Language in Space and Time

In 1916, Ferdinand de Saussure's *Course in General Linguistics* launched structural linguistics and structuralism in social science. A century later, the introduction to linguistics at my own university was titled *Linguistics and Its Structure*. When we think of structures literally, we think of physical things in space, structures like buildings or how bones of an animal fit together. For most of its history, linguistic theory has been expressed via a metaphor of objects arranged in space. We see this repeatedly in many sub-fields of linguistics.

- When writing phonological rules, we say things like one phone is to the left of, the right of, or between other phones.
- An infix is a morpheme inside another morpheme.
- Syllables or phrase structure are commonly expressed via trees where one unit is above, below, or beside another unit.
- Phonology can be said to have layers.
- A language's syntax can be left- or right-branching.
- Within psycholinguistics, we talk about top-down and bottom-up processing.

This predominant use of a spatial metaphor for language becomes surprising when reflecting on natural language, such as a conversation, monologue, or internal thought. Spoken and signed language is ephemeral with little realisation in physical space at all. The physical vibration in space of air molecules dissipates back into the environment incredibly quickly. Fluctuations in the electromagnetic field of our brain or the bodily movement of speech gestures vanish just as fast. While a speaker's tongue certainly occupies physical space, its movement left and right in the mouth does not translate to what linguists mean when phonemes are left or right of each other. Rather than left or right, speech sounds occur *earlier* or *later* than one another.

In short, while linguistic theory is often expressed in terms of a spatial dimension (*in, out, above, below, left, right*), natural language has a strong temporal dimension.

Temporal and spatial dimensions are not mutually exclusive, of course. Language occurs in both. The point here is simply that the temporal dimension of speech is far less researched and considered than a spatial one. This default to spatial language might be because the dominant mode of communicating linguistic theory to a broad audience has been through journals and textbooks, wherein temporally transient language is rendered into semi-permanent spatial configurations via writing. On the page, a symbol indicating a phoneme is indeed to the left of or right of another. On the page with a linguistic tree

structure, a syntactic level is above a morphological one, which is above a phonological one. This is an artefact of writing, of course. When we write using a language's orthography or IPA, a letter (sound) may be to the left of another, but when we speak, the sound is in fact earlier than another. Sounds written on the right are after in time (in left–right orthographies like English). In a spatial tree structure, when X is a node immediately above Y and joined to it, X *contains* Y. In time, Y is inside X when X starts before and ends after Y.

The question for this Element will be: does it matter? Does it matter that something is before something else, rather than to the left/right of it? Or are those equivalent? Does the temporal structure of language affect how we theorise about language? The fact that writing systems can proceed from left to right or right to left or top to bottom, while always being spoken the same way, suggests language is ambivalent about its spatial representation.

This Element narrows the question further. Rather than ask our questions of all of linguistic theory, it will focus only on what is most frequently expressed as 'language at the word level and below'. Expressed in temporal terms, we will focus on processes that happen over about one second of speech, enough for five syllables. In other words, we will look at speech production and perception of words, as well as their phonological patterning. Actions that take longer, such as putting morphemes into phrases, their meaning, their use, and conversation will all be reserved for later work.

The term *dynamic* will be used here in the general sense of 'changing over time', rather than a single approach to the dynamics of language (see also van Geert 1991, 2003). We will sometimes use dynamic systems theory, but this Element works off the principle that, if language changes states over time, it is a dynamic approach to language. The three main aspects of time we will use are:

1. Events can occur earlier, later, or simultaneously with other events. This aspect focuses on temporal sequencing or, using the term most common in psychology, **serial order**.
2. Events have a certain **duration**. They last for a certain amount of time.
3. Events or entities can change over time. This is the study of **dynamics**.

These can be rephrased into a non-exhaustive set of questions that are profitable to ask.

1. How do we produce, perceive, and remember sequences of speech in time?
2. How do we control, produce, and perceive speech durations in time?
3. What are the dynamics of speech?

This Element serves at least two purposes. The first is as a primer to temporal/dynamic approaches to language. The second is to stop and think about

language itself. We want to discover whether asking questions about **time** in language changes how we think of the **structures** of language. In other words, is temporal linguistics an implementation detail of linguistic structure, or does temporal linguistics in fact change structural linguistics itself?

1.2 Who Is This Element For?

The goal of this Element is to introduce people familiar with classic structural linguistics to dynamic approaches to language. By classic structural linguistics, I mean concepts familiar from a year or two of linguistics in most curricula. Sentences are made up of phrases, phrases of words, and words of morphemes. Phonologically, word-sized units contain syllables and syllables contain phonemes. Phonemes in turn can be expressed as contextually dependent allophones, often in a form like:

/phoneme/ ➔ [allophone] / In a particular phonetic environment.

Following such a pattern, the process of vowel nasalisation for /o/ could be written as:

/o/ ➔ [õ] / ____ [nasal]

The reader may know of optimality theory (OT) and how it is a system of constraints that choose an optimal output. This Element will compare its ideas against OT but never engage in doing OT. The following list states what knowledge this Element assumes:

- Introductory articulatory phonetics. You should be comfortable with terms from the IPA like alveolar, palatal, velar, plosives, fricatives, and so on. You will be able to recall their articulatory meanings. For instance, an alveolar involves a tongue tip movement to the alveolar ridge. Similarly, an [s] requires the tongue tip be close but not touch the ridge, while a [t] will have full closure.
- Introductory acoustic phonetics. Familiar terms here would include amplitude, spectrograph, formants, bursts, periodicity, and so on. You will not need to read spectrograms, calculate a Nyquist frequency, or utilise spectral analysis techniques, such as FFT or LPC, but we will utilise spectral analysis conceptually.
- Concepts like phonemes, allophones, distinctive features, syllables, and stress. We will spend much of the Element questioning these very concepts, but these are the starting point. We will also look at metrical phonology wherein languages can assign stress based upon weight, count morae or syllables, and move directionally left to right or right to left (but what does left to right mean if we shift from a spatial metaphor for language to a temporal one?).

- Morphosyntax. This text will not delve into morphosyntax. It only assumes concepts such as morphemes, being the smallest piece of language with meaning, and that morphemes can build words and phrases.

Finally, because this Element is part of the Psycholinguistics series from Cambridge University Press, it assumes you are familiar with common questions of psycholinguistics and some of the techniques. You would know what speech perception, speech production, and word recognition are, as well as generally how these questions are studied. However, I do not expect the reader to have Levelt's (1993) model of speech production in mind, be able to recall the details of TRACE (McClelland & Elman, 1986), and so on.

This Element's assumptions about your knowledge are based on the purposes of the text. One reason dynamic approaches have not received as much attention as they warrant is they can be challenging for a linguist of typical training. Just as it would be difficult to pick up an article deriving prosodic patterns via OT when you have not studied structural linguistics, research on dynamic systems approaches can challenge linguists. It can be easy to get lost in the technical terminology of attractors, limit cycles, evolving systems, differential equations, and more. *Dynamic Approaches to Phonological Processing*[1] hopes to build a bridge from structural linguistics to this dynamic literature. To accomplish this:

- This Element simplifies models. Sub-components are sometimes left out or technical details are passed over. As one example, Turk and Shattuck-Hufnagel's (2020) book *Speech Timing: Implications for Theories of Phonology, Phonetics, and Speech Motor Control* covers its ideas over 300 pages. We will cover their theory in fewer than 5 pages. Every theory presented here is in fact richer, more substantiated, and precise than this work has space for. I hope this Element can prepare the reader to approach the original work for full engagement.
- The Element offers a select set of dynamic models and does not attempt a comprehensive look as a review article might. Covering such breadth would sacrifice the space for introducing concepts at the right introductory level or going deep enough into some to highlight their potential value.

2 Serial Order

Any complete model of language will eventually need to specify how things are put into and kept in the right order. In a spoken conversation (as opposed to

[1] Phonetic transcriptions by default will be in New Zealand English, based on Bauer and colleagues' (2007) analysis. Exceptions include using transcriptions from research and a couple of points where alternate transcription allows for easier explication.

written text), order requires saying some items before or after other items, which means taking certain actions at certain times. Therefore, serial order is inherently about controlling when to do something. The items to order might be phrases, words, morphemes, syllables, phones, all the way to gestures of the hands or vocal tract. If we are listening, we also need to process the units of language as they arrive in some order as well. Was this acoustic burst, phone, word, phrase, or clause first or second? Even if incoming material is processed immediately, any ability to reflect on what was previously said requires knowing what 'previous' is.

The question of serial order is hiding in every IPA transcription or statement that a certain series of phonemes is the phonemic representation of a word. As Kazanina, Bowers, and Idsardi (2018) state,

> A traditional answer from linguistic theory . . . is that words are represented in long-term memory as sequences of phonemes, that is, abstract and discrete symbolic units of a size of an individual speech segment, such as a consonants or vowel (yet not identical to them). *A phonological form of a word is an ordered sequence* of phonemes, for example, the sequence of phonemes /k/ – /æ/ – /t/ (more succinctly, /kæt/) refers to a meowing domesticated feline animal or /d^k/ to a quacking avian. (561; emphasis added)

In short, every single word with a phonological representation requires learning, maintaining, and producing a serial order, a sequence of actions in time. The phonological form of a word is, therefore, two things: a set of phonemes and an order of those phonemes. Phonemes are categorical units capable of building lexical contrasts, but a set of phonemes is not sufficient to represent a lexical item: the phonemes' order must also be specified and remembered. We would not understand a lexical representation without understanding serial order. How one maintains serial order for phonemic representations is not commonly addressed in linguistics textbooks – but we will do so now.

Two seminal works from the 1950s have guided the study of serial order and serial recall. One is Lashley's (1951) *The Problem of Serial Order in Behavior*, which documented how turning hierarchical mental structures into serial actions was a fundamental question across a range of behaviours. The second seminal work is Miller's (1956) 'The magical number seven, plus or minus two: Some limits on our capacity for processing information'. Miller's article claimed humans can hold about seven items in what came to be called working memory, based primarily upon research in the serial recall experimental task. In such an experiment, a participant is presented with a list of items, usually short words, and must repeat them in order. For example, the participant hears *nine, four, six, two, one, three*, and attempts to repeat this sequence without error. It is

noteworthy for us that the classic paradigm of this experiment is a linguistic task: the participant perceives speech (or reads it) and then produces speech in turn. Serial recall is a controlled linguistic exercise.

Serial recall has been extensively researched over the last seventy years, and a model of serial recall and serial order today must meet a large number of empirical constraints to be a viable model. Because many factors have been shown to affect serial recall performance, detecting a new one requires very careful stimulus construction and control (for a recent review, see Hitch, Hurlstone, & Hartley, 2022). For this work, we will skip directly towards influential models that will push our discussion of time in language forward.

2.1 Competitive Queuing

One key mechanism used in a variety of serial order models is competitive queuing (Bohland, Bullock, & Guenther, 2010; Burgess & Hitch, 1992, 1999, 2006; Grossberg, 1978; Houghton & Hartley, 1995; Lewandowsky & Farrell, 2008). With competitive queuing, several items are all activated in parallel and then 'compete' to be selected. Competition includes each item exciting itself – trying to increase its own activation – and inhibiting its competitors – lowering their activations. Which one will win? In general, the winning candidate will be the one that is most active at the start, because that one has more energy to outcompete the other candidates. The leading candidate's own activity increases while it inhibits its competitors' activity. When the competitor reaches a threshold, the item is **selected** and initiates. For instance, if the competing items were a set of words, then one word would win and be uttered.

However, if the process stopped here, then the winner would be selected in perpetuity, endlessly being re-selected. To say the second word, which had a lower starting activation than the first, we must get the first one out of its way. This happens through a **suppression** process. After an item passes the threshold, a feedback mechanism strongly suppresses its activation so that it plummets to a level smaller than the competitors. This gives space for the second most active item to be selected. This process repeats as long as there are active items (Figure 1).

Competitive queuing then could be a deterministic process: the order of selection is determined entirely by the activation levels of the items at each time point. Noise in activation levels can disrupt this, however. If the noise, or randomness, is very low, then almost every selection will be determined by the specified activity levels at the start. With increased noise, it becomes possible for a less active candidate to get a random boost over a more active one.

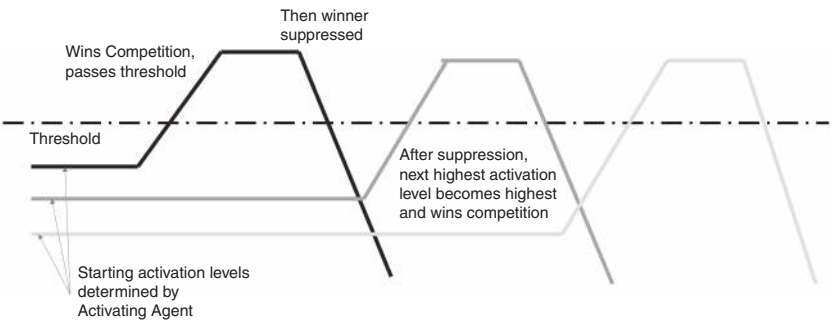

Figure 1 Schematic representation of three items in competitive queuing

Noise then becomes a source of error. If the goal of the competitive queuing model is to model human behaviour, this error becomes a virtue. Errors are a primary guide to discover the sort of factors that guide a speaker's mental models. For instance, errors that switch out the beginnings of words can be evidence that beginnings are a constituent, to use a linguistic concept. Similarly, more errors on items in the beginning of a sequence than on the first item in a sequence gives further evidence of how our mind is handling sequences. A strong psychological model generates the sorts of errors that humans in fact make and avoids generating errors they do not make (Hurlstone, 2021). With competitive queuing, we have a system where activation levels of competing items determine what will be selected. What determines those activation levels, though? The activating agent.

To understand this, it might be best to step back and recall the goal. We want to build a model of how a speaker can hear items in a sequence, remember the items and the sequence, and then recall both when prompted. To give a concrete example, they might hear a sequence of numbers:

9 2 8 5 1 3

Two main types of mistake occur when repeating this sequence. The speaker might err about what numbers were said, such as saying '7 2 8 5 1 3' when no 7 was present in the heard sequence. This would be an **item error**. The speaker also might make an **order error**, such as saying, '2 9 8 5 1 3', which switches the first two items in the sequence but includes all the correct items. Therefore, models have two primary components.

Component One represents the items. This could be a simple representation like using the exact number symbols, but it could be more complex as well. To take an example, let us say we want to create an item representation for the word *spoon*. We could use phonemic symbols:

s p ʉː n

The entire word is represented with a vector of four dimensions, one dimension for each phoneme. A limitation to this representation is that it has no notion of phonetic similarity. This does not match empirical evidence. *Spoon* and *spool* are confused more often than *spoon* and *cheese*. To model facts such as this, researchers can introduce a phonetic characterisation into the item representation. One solution would be to break each phoneme apart into pseudo-features:

s ➔	alveolar	voiceless	frication	unrounded
p ➔	bilabial	voiceless	plosive	rounded
ʉː ➔	palatal	voiced	vowel	rounded
n ➔	alveolar	voiced	nasal	unrounded

This represents each sound with a four-dimensional vector (which is of course partial), needing the full 4x4 array to represent a word. With this more complex representation, sounds that share more features would be confused more often than sounds that share few features. If a model with the posited phonetic representation produces results similar to actual speakers, then it is taken as evidence that people might indeed use that representation.

Component Two is a representation of the order, often called the context or learning context because it is the context in which the items are heard. For a repetition experiment, we want to store the context of the items and then repeat those items based upon the context. Therefore, it is this **context signal**, which is the activating agent. The context signal will set the activation levels for the items for competitive queuing. Many different models therefore can use competitive queuing as the selection mechanism but change the context signal to better match empirical results of speaker behaviour. In our next sections, we will examine a model that uses a dynamic system in the form of cascading oscillators to represent the context signal. To deal with this model, we first need to look at what a dynamic system and oscillator are.

2.2 Dynamic Systems and Oscillators

A dynamic system is any system that changes states over time (Figure 2). A traffic signal can be interpreted as a dynamic system with three states. At time t_1, it has a state of green; at time t_2, a state of yellow; and at time t_3 a state of red.[2] Green, yellow, and red are its **states** or **state space**.

[2] We are ignoring other states such as flashing versions of the colour. Moves back to a colour in a sequence do not change what the possible states are, only the sequence of those states.