Dickerson, W. B. (2020). Deciphering everyday speech. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, Northern Arizona University, September 2019 (pp. 67–75). Ames, IA: Iowa State University.

DECIPHERING EVERYDAY SPEECH

Wayne B. Dickerson, University of Illinois at Urbana-Champaign

Language learners' frustration when first trying to understand native speakers' casual speech is so common that many consider it an expected part of language acquisition. However, research suggests that encounters with everyday speech need not be so discouraging if ESL/EFL pronunciation instructors would prepare learners for the experience. This paper argues that listening instruction would be more effective if instructors began with a more accurate conception of what spontaneous speech is really like and the nature of the rules that describe it.

INTRODUCTION

Pronunciation teachers are quite fond of rules. When asked to describe how native English speakers would pronounce the phrase *So I am going to take Japanese,* we do not hesitate to say that the words *so* and *I* will be linked together by the back-rounded offglide of the vowel in *so*, that in contracting *am* we drop its vowel, that *going to* will be pronounced as [gənə], and that the final consonant of *take* will link to the initial consonant of *Japanese* as an unreleased [k]. We might add that our students can make all of these predictions themselves because we prioritize connected-speech rules in class.

If we are candid, however, we have to concede that, even after practicing their streamlining skills, learners may still fail to understand the very same string—*so I'm going to*—in conversations. Even though they are prepared to hear [səʊaɪmgənə], they are unprepared to hear [swɑməm̃ə] that has no vowel nucleus for *so,* uses its remaining back-rounded offglide to create an [sw] cluster, drops the front-unrounded offglide after the nucleus of *I*, drops the [g] of *gonna*, and taps the final nasal consonant [m̃]. In the last two words, *take Japanese,* the [k] of *take* may even change to a glottal stop [ʔ], and *Japanese* may lose its middle syllable while the [p] becomes unreleased before the [n]. Such deviations from what learners expect are enough to render this phrase unintelligible to their untrained ears (Munro & Derwing, 1995).

Wherever learners study English, their earliest encounters with spontaneous English speech are much the same. At the start of my pedagogical phonology course, I regularly ask new MATESOL candidates to relate a memorable story from their language-learning experience. One of the most frequent accounts, mainly from international students, describes their shock of not understanding questions or answers on first arriving in the States. For some, the discomfort of that experience remains as vivid as yesterday.

This disappointing scenario is so widespread that we think of it as part of the usual progression toward language acquisition. This paper argues that the likely culprit is an ineffective pedagogical approach to deciphering conversational speech. It arises from ESL/EFL teachers' long-standing misconception about the nature of language rules. The misunderstanding colors our orientation to the problem and limits how we prepare learners for everyday social interactions.

PRONUNCIATION RULES

These days we embark on pronunciation instruction with a toolkit of rules to bring predictability to the English sound system and a sense of control to learners. We tell our students, for example:

(1) A word-final [t] before the [j] of *you, your, yourself, yet,* and *year* is pronounced [ʧ], as in *Is this wha**t you** want? Sui**t yourself**!*
(2) Compound nouns, like *staff meeting* and *soccer field,* carry heavy stress on the first word.
(3) The main prominence of a phrase goes on its last content word.

Teachers may not realize that before the 1950s, few pronunciation texts had any rules (Pike, 1942; Clarey & Dixson, 1947). Clifford Prator, Jr. published the first rule-based text in 1951. Thanks to him, we expect rules for all the topics in our syllabus and would feel at a loss without them.

The Authority of Rules

Rules definitely have, and should have, a place in pronunciation instruction. However, a sober assessment of their largely unnoticed liabilities might temper our enthusiasm for the rules we teach.

First, what we call rules in English pronunciation are best viewed as tendencies. At worst, they may simply be wrong. Despite the connotation of the word *rule* as mandated behavior, no pronunciation rule carries such weight. All are approximations of what we hear through biased ears made deaf by familiarity; they all misrepresent English to some degree and merit our skepticism. As we simplify our rules for learners, we also inevitably introduce distortions. We would do well to use words like *often, usually,* and *almost always* with our rules to be explicit about the limits of our understanding and the extent of their authority. To his credit, Prator states his rules with some leeway: "Accents tend to recur at regular intervals. … In general content words are stressed" (1951, p. 25).

Rules of connected speech should be regarded with the same caution. In my teaching, I refer to them as "naturals"—N-A-T-R-L—Native Assimilation, Trimming, Reduction, and Linking.[i] However, I emphasize that they only scratch the surface of the changes that occur in everyday speech.

A second concern is that time and repetition can endow rules with the authority of truth, which can short circuit our pursuit of accuracy. When we take rules as statements of fact, we no longer keep up with developments in the area of the rules. An extreme case of failing to exercise due diligence relates to Prator's two rules (1951, p. 25)—or "principles" as he called them—regarding stress timing and accenting content words, cited above. We asserted their accuracy for decades after they were shown to be unsupported (Cauldwell, 2002). Rules of connected speech are also prone to be neglected. When we believe they fully describe naturally occurring speech, it is easy to assume that nothing more can be learned about them. This attitude can arrest our interest in exploring further and limit the possibility of new discoveries.

Some rules promote the harmful tendencies just expressed—overstating our knowledge about language structure which can, in time, undermine our curiosity to investigate it further. Other rules

can inoculate us against these tendencies. Two types of rules are involved: categorical and variable.

The Nature of Rules: Categorical and Variable

Another connotation of the word *rule* is that it applies always and everywhere with the same result. A rule of this type is called a **categorical** rule. Rules (1)-(3) above are examples of categorical rules.

In some areas of experience, categorical rules are appropriate, such as the rules of the road we study to pass a driving test in a particular state or country. They regulate the safe movement of traffic and are the basis for the traffic tickets we receive when we violate the rules.

In other areas, categorical rules have no place, as when describing the sound system of a language. It is not static in form but inherently full of variability because it is constantly being changed by those who speak it. To describe the sound system requires the kind of rule that reflects its inconsistencies, namely, a **variable rule**.[ii] A variable rule sometimes applies and sometimes not, or applies only a portion of the time. For example, a more faithful statement of rule (1) above, including its variability, is the following.

A word-final [t] before the [j] of *you, your, yourself, yet,* and *year* becomes [ʧ], [tʃ], [ʃ], [ʔʃ], or [ʔj] (Shockey, 2003, p. 38, 44f), as in *It won't hur**t you*** [hɝʧə] or [hɝtʃə], *You can le**t yourself** in* [lɛʃɚsɛɤf], *Star**t your** car* [stɑrʔʃɚ], and *No**t yet*** [nɑʔjɛt].

Variable rules allow more refined descriptions of English. For example, variants in the output of a variable rule may not be in free variation. Instead, the proportion of each variant might characterize different styles of speech of different age groups in different social or ethnic groups. Showing patterns in variability is the kind of work that sociolinguists like William Labov (1972) do. Variable rules can also reveal patterns of change in second-language acquisition (Dickerson, 1976).

Rule types and their susceptibility to mishandling should now be clear. Categorical rules are liable to misuse because they make unequivocal claims about their unchanging output. Variable rules caution us against such misuse by keeping the range of their variable outputs in front of us.

Given how thoroughly we are schooled in the categorical nature of pronunciation rules, most teachers are unaware of how much classroom talk differs from everyday conversational speech because of the oversimplifications of English found in our textbooks. Given our blindness to the phonetic reality of our speech in unrehearsed interaction, it is no wonder we are suspicious of variable rules. A common reaction is to be dismissive: Why go into such detail? Isn't the big picture enough for learners? This reaction may arise from a concern not to overwhelm learners with minutiae and what may seem like trivial, inconsequential details. One acceptable pronunciation is no doubt good enough when teaching learners to produce, but it is certainly inadequate preparation for deciphering the range of variation reaching their ears in casual interactions.

Another equally common reaction is disbelief. Demonstrations of compressed speech—what Cauldwell (2018) calls "the rags and shreds of words mushed up into an acoustic blur" (p. 23) — can be startling and difficult to accept as a personal reality: I don't sound like that, do I?[iii] This

very naïveté on our part and our reluctance to engage with phonetic content that seems unrepresentative if not grossly improper are some of the reasons our students cannot handle spontaneous speech outside the classroom.

The difference between categorical and variable rules is likely to be missed in Cauldwell's (2018) portrayal of spontaneous speech because he concludes that "it is probably best to regard it as a domain … where there are no rules" (p. 44) and therefore can be justifiably characterized as *unruly* (pp. 13-15, 36). While *unruly* fits his botanical metaphor, in which he likens such speech to a jungle, it does not fit the reality of jungle speech. It is true that jungle speech cannot be described adequately by categorical rules, but it is not unruly in the sense that it follows no rules. Rather, it follows variable rules which both Cauldwell (2018, p. 32, 83ff) and Shockey (2003, 14ff) call "streamlining processes" or just "processes." Both go to considerable lengths to spell out the environments in which a relatively small set of processes occur variably.

THE IMPORTANCE OF VARIABILITY TO INTELLIGIBILITY

The messy phonetic reality of spontaneous speech is far too important to dismiss or disbelieve. To understand its centrality to all oral communication, we must recognize its essential function.

In the era of stress-timed rhythm, we believed that we used valley compression to promote the regular occurrence of accents (Prator, 1951, p. 25). We now know that English rhythm is not stress-timed (Cauldwell, 2002). That means we compress our words with some other intent. The reason we streamline parts of phrases, fragmenting and discarding sounds as we go, becomes clear in light of a different model of rhythm, a particular way of packaging meaning for listeners, and the exigencies of through-the-air communication.

Briefly, the picture is as follows. Spontaneous speech is characterized by short bursts of language, nearly always with only one or two prominent syllables or pitch accents (Bolinger, 1961; Brazil et al., 1980). Its rhythm is not in its timing of pitch accents but in the alternation between accented and unaccented parts (Cauldwell, 2002). Semantically, the most salient parts of these phrases are either the single prominence (and surrounding function words) when there is only one prominent syllable or the stretch of language from the first through the second prominence when there are two prominent syllables (Brazil et al., 1980, pp. 43, 45; Wells, 2006, pp. 233-234). Speakers shorten unaccented stretches of a phrase by shedding sounds and syllables as they speak. In this way, they move to and through the salient parts more quickly, not necessarily by speaking faster, but by saying fewer sounds and syllables (Shockey, 2003, pp. 11-13). These abbreviated strings help listeners, who have limited short-term memories and processing capacities, to snatch the full essence of each phrase out of the air in the milliseconds of its existence (Kjellin, 1999). The better that speakers can deliver an intact semantic unit to listeners' ears, the better that listeners will understand it as a single thought and remember its message (Hahn, 2004).

The opposite effect also holds. Long phrases, slow speech, more than two pitch accents, non-essential sounds, and extraneous pauses increase time-per-phrase and challenge listeners' short-term memory (Hahn, 2004; Levis, 2018, pp. 26-27). The cost to listeners can be the loss of intelligibility. Failing to understand the core meaning of a phrase in one take, listeners often turn their attention to figuring it out, potentially missing the speaker's upcoming phrases altogether.

The graphic we use in pronunciation instruction to depict this rhythm is the **two-peak profile** in Figure 1. In the metaphor of a mountain range in silhouette, the second (or only) pitch accent is the **primary peak** (●) and the first, when there are two pitch accents, is the **anchor peak** (○). The syllables surrounding these peaks we call **valleys,** where all the streamlining occurs.

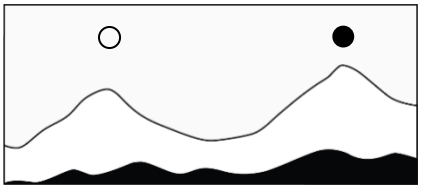

Figure 1. The Two-Peak Profile.

The variability that results from streamlining speech is functional in, and integral to, oral communication. To ignore or dismiss this variability is to perpetuate our present predicament. Ideally, we should focus on variability where learners need it most. To the credit of TESOL professionals, pronunciation instruction that includes categorical connected-speech rules does prepare ESL/EFL learners to deliver intelligible speech to listeners (Cauldwell, 2018, pp. 61, 70). Learners need not aspire to **produce** native-like compression found in spontaneous speech.

By contrast, learners who hope to hold up their end in spontaneous conversations do need to decode and interpret the compressed speech they hear. Listening practice which is limited to the connected-speech rules we teach in class is inadequate preparation. The differences between [səʊaɪmgʌnə] and [swɑməʔə] reveals much more streamlining and variation in unrehearsed talk than we cover in class. The gap in our instruction in the area of **perception** is so strategic for understanding extemporaneous talk that learners need dedicated help to decode the variability in this kind of talk. This is the asymmetry of production and perception that instructors would do well to address (Levis, 2018, p. 148).

IMPLEMENTING THE ASYMMETRY OF PRODUCTION AND PERCEPTION

Despite their limitations, categorical connected-speech rules are still needed. Besides yielding intelligible speech, they also help learners acknowledge the reality of variability.

- Compared to words pronounced one-by-one, connected-speech rules suggest that words have multiple pronunciations particularly at their edges. Understanding this fact makes it easier for learners to accept the variable pronunciations of spontaneous speech.
- Even though categorical connected-speech rules are considered part of careful speech, they nevertheless hint at spontaneous speech. It is easier to present the variability of casual speech to learners when some of the steps in that direction are already familiar.
- Native speakers compress their speech to make it easier for listeners to grasp the core message of each phrase. Knowing this can motivate learners to take advantage of the help that speakers offer. That, too, prepares learners to take the next critical step in deciphering.

Therefore, for purposes of production, we teach the traditional, categorical rules of connected speech—the "naturals"—but acknowledge their limitations in describing all sounds in spontaneous speech.

For purposes of perception, Cauldwell (2018), Field (2003), and others suggest how to prepare students to decipher casual, everyday speech. An examination of various proposals to implement decoding training reveals that, at root, training involves three steps. The following are the steps that Laura Hahn and I have incorporated into every lesson of *Speechcraft: Discourse Pronunciation for Academic Communication,* 2nd edition (in press).

1. Target learners' greatest listening problem: word clusters in spontaneous-speech valleys. Word clusters—also called "lexical bundles" (Biber et al., 2004)—are high-frequency groups of two, three, or four non-prominent words. They consist largely of function words. An example of a word cluster is *so I'm going to,* as in *So I'm going to take Japanese.* Cauldwell provides other examples in an appendix (2018, pp. 230-232). The recommendation is to always attach a word cluster to a pitch-accented word (Cauldwell, 2018, p. 31). In this case, the word cluster is attached to the word *take* (the anchor peak).

2. Expose learners to a variety of phonetic forms of each word cluster in order to develop their familiarity with the underlying processes. A word cluster has no single pronunciation in our personal usage, much less from person to person. It is impractical in most classroom settings to cover more than a fraction of possible word clusters. However, the few streamlining processes that underlie the variability of word clusters offer a reasonable target. This is why, for *Speechcraft* lessons, we have selected word clusters to expose learners repeatedly to the variety of processes at work in spontaneous speech.

3. Use pronunciation exercises to help learners perceive word clusters and associate meanings with them, but not for the purpose of teaching learners to pronounce them. Experience teaching phonetics confirms that learning to say unfamiliar sounds can promote better perception of those sounds (Levis, 2018, p. 149).

Only 4-5 minutes are needed to step through this sequence to introduce a deciphering exercise in class. The result of repeating this type of exercise lesson after lesson is that learners gradually improve their ability to recognize sound shapes of word clusters by ear and associate meanings with them. They also become familiar with the small number of underlying processes responsible for these sound shapes. This proposal is intended to supplement whatever phrase-level listening instruction, if any, is currently being used.

A SAMPLE INTRODUCTION TO DECIPHERING

The following is an example of a four-minute, in-class exercise of the type we use in *Speechcraft* to introduce a word cluster. We call the exercise *Deciphering Spontaneous Speech Sounds.* Its intent is to go beyond "naturals"—categorical connected-speech rules—in order to illustrate the effects of variable rules on valley syllables spoken spontaneously. Instructions for teaching each deciphering activity are detailed in the instructor guide accompanying *Speechcraft.*

Deciphering exercises begin with the first substantive *Speechcraft* lesson for two reasons. We want to build up an expectation that perception practice for decoding is an integral part of pronunciation practice. Furthermore, this exercise makes explicit that spontaneous speech, not careful speech, is the ultimate goal of our perception work.

To connect sound to meaning, we draw phrases from the content of the lesson itself. This example would come in the third lesson. We begin by saying something to this effect to the class:

> "We've just used the phrase, *So I'm going to take Japanese.* But when you leave class, and the expression, *so I'm going to,* comes up in conversation, it may sound different from the careful way we pronounce it in class, either *so I'm going to* or *so I'm gonna*. The careful version starts you toward hearing spontaneous speech. This exercise takes you further. It introduces some of the ways speakers pronounce the phrase in conversations you may have. You'll practice listening to and repeating these different pronunciations at home, using audio recordings. This will help you recognize the phrases when you hear them in spontaneous speech. Let's begin with a word-by-word version of the phrase to be sure you understand every word."

After saying the phrase, *so-I'm-going-to-take-Japanese,* and asking for questions about its meaning, we move to the careful-speech version. For this, we explicitly recognize all "natural" rules studied to this point, and often solicit them from class members, such as the pronunciation of *going to* as *gonna*. We ask for a few repetitions of this version of the phrase before turning to spontaneous versions.

We introduce conversational versions in each lesson by repeating that (a) practice producing these streamlined versions has been shown to help perception; we do not expect learners to use the forms in their own speech because careful speech is clear enough to listeners, and (b) despite the changes to valley syllables, speakers often leave traces of the original words so that listeners will know what the valley words are. Then we identify each change affecting each spontaneous-speech version but do not dwell on the process itself. Too much detail tends to break the rhythm of the exercise and to overload learners' capacity to absorb new information when we introduce variants. Next, for each version, we ask learners to repeat the target word cluster with its accented word (underlined below) twice (x 2) and then the whole phrase twice (+ 2). We do everything orally so that symbols do not interfere with the process.

First, the off-glide of *I* [aɪ] is trimmed.	[səʊamgənəˌtʰeɪkˀʤæpəˈniːz]	x 2 + 2
Next, the en of *gonna* is tapped.	[səʊamgə̃ɾəˌtʰeɪkˀʤæpəˈniːz]	x 2 + 2
Sometimes we hear the em of *am* as eng before the [g] of *gonna.* At other times, the gee of *gonna* drops out.	[səʊaŋgə̃ɾəˌtʰeɪkˀʤæpəˈniːz]	x 2 + 2
	[səʊamə̃ɾəˌtʰeɪkˀʤæpəˈniːz]	x 2 + 2
The nucleus of *so* [əʊ] can drop to create an [sw] cluster.	[swamə̃ɾəˌtʰeɪkˀʤæpəˈniːz]	x 2 + 2

For homework, we assign the associated audio recording. The bulk of students' practice time is devoted to simultaneous production, namely, repeating the phrase **with** the model 20-30 times by putting the recording into loop mode (Kjellin, 1999).[iv]

When students return to class, we review what they have practiced. Then we check their perception of the target word cluster. After we pronounce a phrase like the following examples with any one of the variants students have practiced, we ask them to write what they hear and then, individually, to read in careful speech what they wrote. This is one way to confirm their comprehension.

So I'm going to take chémistry.	So I'm going to wárn them.
So I'm going to take Árabic.	So I'm going to vísit him.
So I'm going to take philósophy.	So I'm going to búy some.
So I'm going to take phýsics.	So I'm going to téll you.

CONCLUSION

The mismatch between how we describe connected speech in ESL/EFL materials and the reality of spontaneous speech can be summed up as the difference in the output of categorical and variable rules. Because the phonetic diversity in spontaneous speech can prevent learners from understanding their co-speakers, our obligation as instructors is to give them decoding practice to familiarize them with the processes underlying this diversity. Variation in connected speech, far from being superfluous, is the key to learners' understanding such speech. The expansion of teaching content to enable them to interpret this phonetic variability—only 4-5 minutes of each class—is meager compared to its potential to improve learners' skill to decipher everyday speech.

ABOUT THE AUTHOR

Wayne Dickerson is professor emeritus in the Department of Linguistics at the University of Illinois at Urbana-Champaign where he directed the MATESOL program and taught courses in English phonology and ESL pronunciation. His two pronunciation textbooks are *Stress in the Speech Stream: The Rhythm of Spoken English* (1989/2004), University of Illinois Press, and (with co-author Laura D. Hahn) *Speechcraft: Discourse Pronunciation for Advanced Learners* (1999), University of Michigan Press.

Contact information:
Wayne Dickerson
7 Hale Haven Court
Savoy, IL 61874 U.S.A.
dickrson@illinois.edu

REFERENCES

Biber, D., Conrad, S., & Cortes, V. (2004). If you look … Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.

Bolinger, D. (1961). Three analogies. *Hispanica, 44,* 135-136.

Brazil, D., Coulthard, M., & Johns, C. (1980). *Discourse intonation and language teaching.* London: Longman.

Cauldwell, R. (2002). The functional irrhythmicality of spontaneous speech: A discourse view of speech rhythms. *Apples, 2,* 1-24.

Cauldwell, R. (2018). *A syllabus for listening, decoding*. Birmingham, UK: Speech in Action.

Clarey, M., & Dixson, R. (1947). *Pronunciation exercises in English.* New York: Simon & Schuster, Inc.

Dickerson, W. (1976). The psycholinguistic unity of language learning and language change. *Language Learning, 26*(2), 215-231.

Dickerson, W., & Hahn, L. (in press). *Speechcraft: Discourse pronunciation for academic aommunication* (2nd ed.). Ann Arbor: University of Michigan Press.

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal, 57*(4), 325-334.

Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*(2), 201-223.

Kjellin, O. (1999). Accent addition: Prosody and perception facilitates second language learning. In O. Fujimura, B. Joseph, & B. Palek (Eds.). *Proceedings of LP '98 (Linguistics and Phonetics Conference),* Ohio State University, Columbus, OH (pp. 373-398). Prague: The Karolinum Press.

Labov, W. (1966). *Social stratification of New York City.* Washington, DC: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic patterns.* (Conduct and Communication, 4.) Philadelphia: University of Pennsylvania Press.

Levis, J. (2018). *Intelligibility, oral communication, and the teaching of pronunciation.* Cambridge, UK: Cambridge University Press.

Munro, M., & Derwing, T. (1995) Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 48*(2), 159-82.

Pike, K. (1942). Pronunciation. In *Intensive course in English for Latin-American students*. Volume 1. Ann Arbor: The University of Michigan Press.

Prator, C., Jr. (1951). *Manual of American English pronunciation for adult foreign students.* Los Angeles: University of California Press.

Shockey, L. (2003). *Sound patterns of spoken English.* Oxford, UK: Blackwell Publishing, Ltd.

Wells, J. (2006). *English intonation, an introduction.* Cambridge, UK: Cambridge University Press.

[i] **A**ssimilation includes palatal and voice assimilation; **t**rimming includes contraction, cluster simplification, f-elision, h-elision, vowel elision without syllabics; **r**eduction includes vowel reduction and types of consonant reduction such as tapping; **l**inking includes consonant-to-vowel linking, vowel-to-vowel linking, consonant-to-identical consonant linking, and stop-to-(stop, affricate, nasal) linking.

[ii] The introduction of variable rules by Labov (1966), the father of variationist sociolinguistics, contradicted the categorical rules of generative phonologists at the time. Labov showed that some of the variability, swept under the rug of performance (of the competence-performance dichotomy) as theoretically uninteresting, is critical evidence of how language changes over time.

[iii] In our pedagogical phonology course for MATESOL candidates, we identify personal tendencies that prevent our hearing speech as it truly is: hypercorrection, spelling pronunciations, the observer's paradox (Labov, 1972), and a general unawareness of how we actually sound. By directly addressing MATESOL students' resistance to accept variation in their own speech, we dramatically reduce their skepticism.

[iv] Students download and use the free, high-quality audio playback, recording, and editing program, *Audacity*.