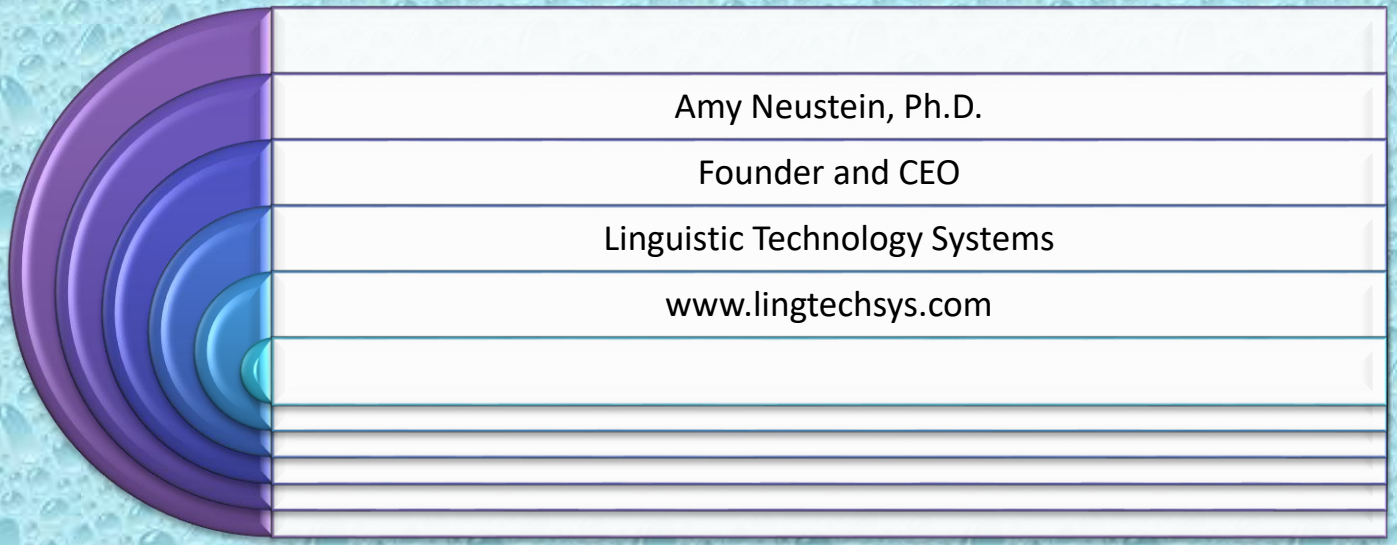
International Conference on Soft Computing Models in Industrial and Environmental Applications

University of Salamanca

April 6-8, 2011



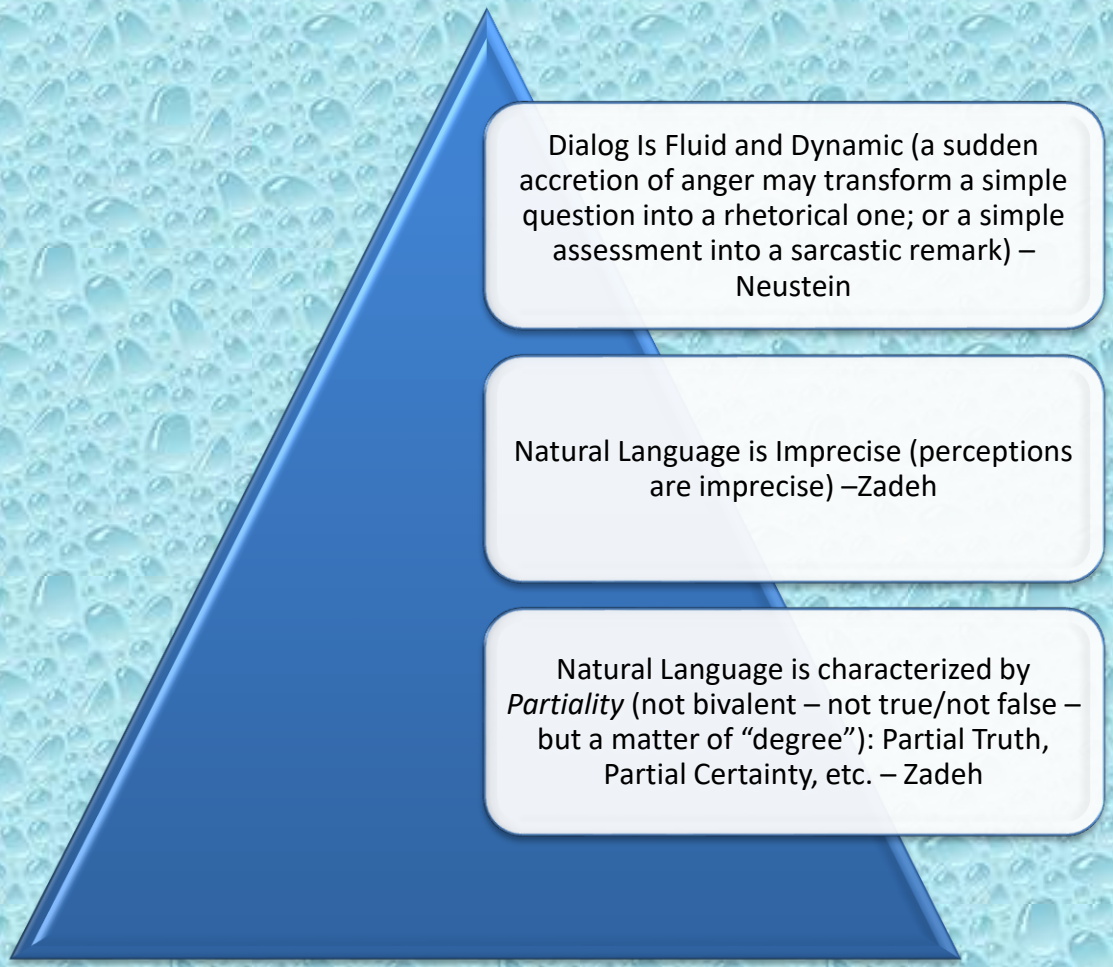
Amy Neustein, Ph.D.

Founder and CEO

Linguistic Technology Systems

www.lingtechsys.com

Sequence Package Analysis and Soft Computing
Introducing a New Hybrid Method to Adjust to the Fluid and Dynamic Nature of Human Speech



What Does Sequence Package Analysis (SPA) Have in common with Precisiated Natural Language (PNL)?

Both SPA and PNL approaches to natural language abandon bivalent logic – a logic in which shading of or "degree" of truth is not allowed;

Both SPA and PNL exploit the imprecision of natural language to transform it into a precise formal construct;

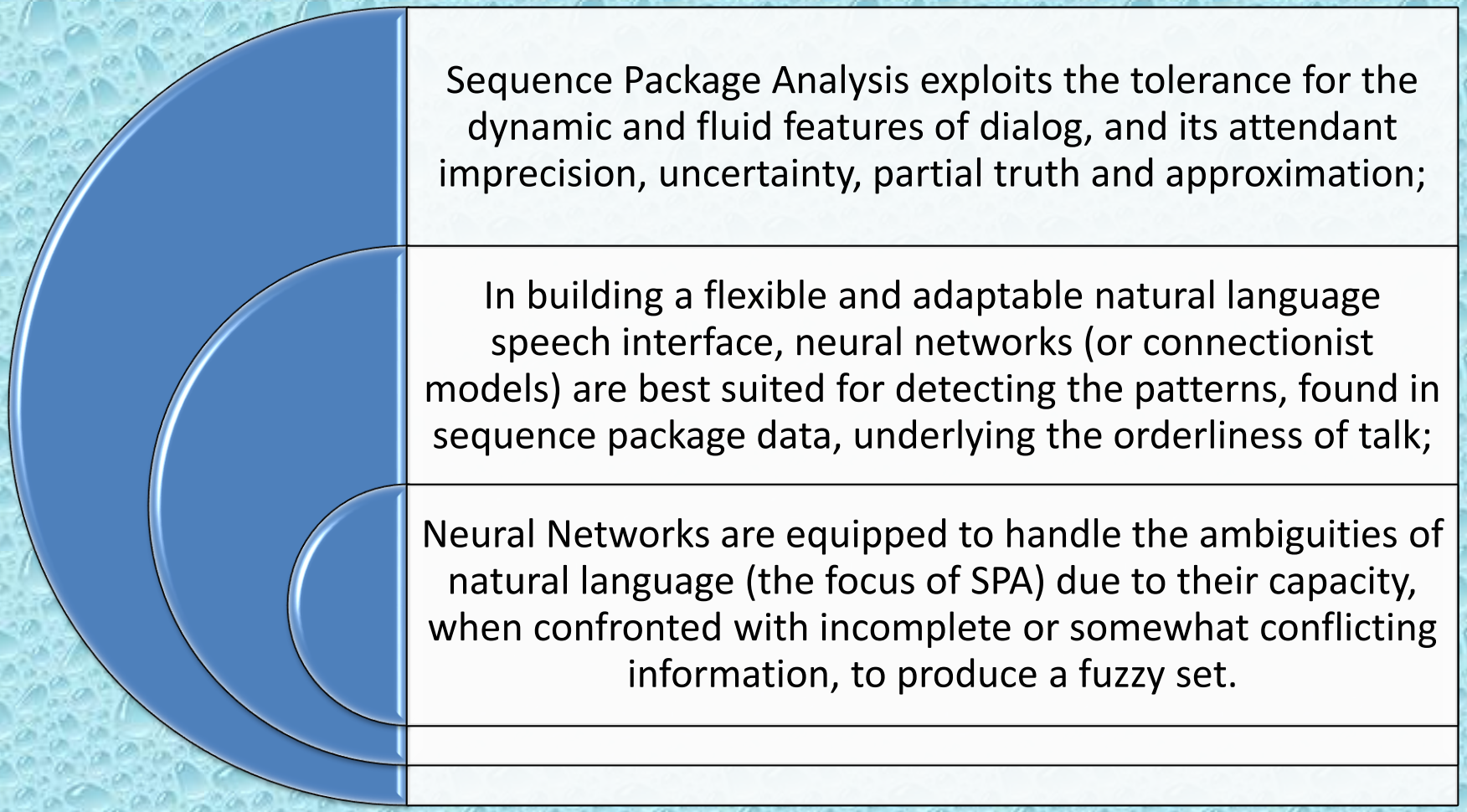
Both SPA and PNL achieve tractability, robustness and low solution cost in real-world problems.

TRANSFORMING IMPRECISION INTO FORMAL CONSTRUCTS

In PNL, precisiation is accomplished through translation into what is termed a precisiation language: known as a generalized-constraint language (GCL) – a language whose elements are so-called generalized constraints and their combinations (Zadeh);

In SPA, precision is achieved through the interpolation of sequence package information – using a novel POS (part of speech) tagging program – into the speech engine's output stream (Neustein).

Why is SPA a Complementary Hybrid Method in Soft Computing?

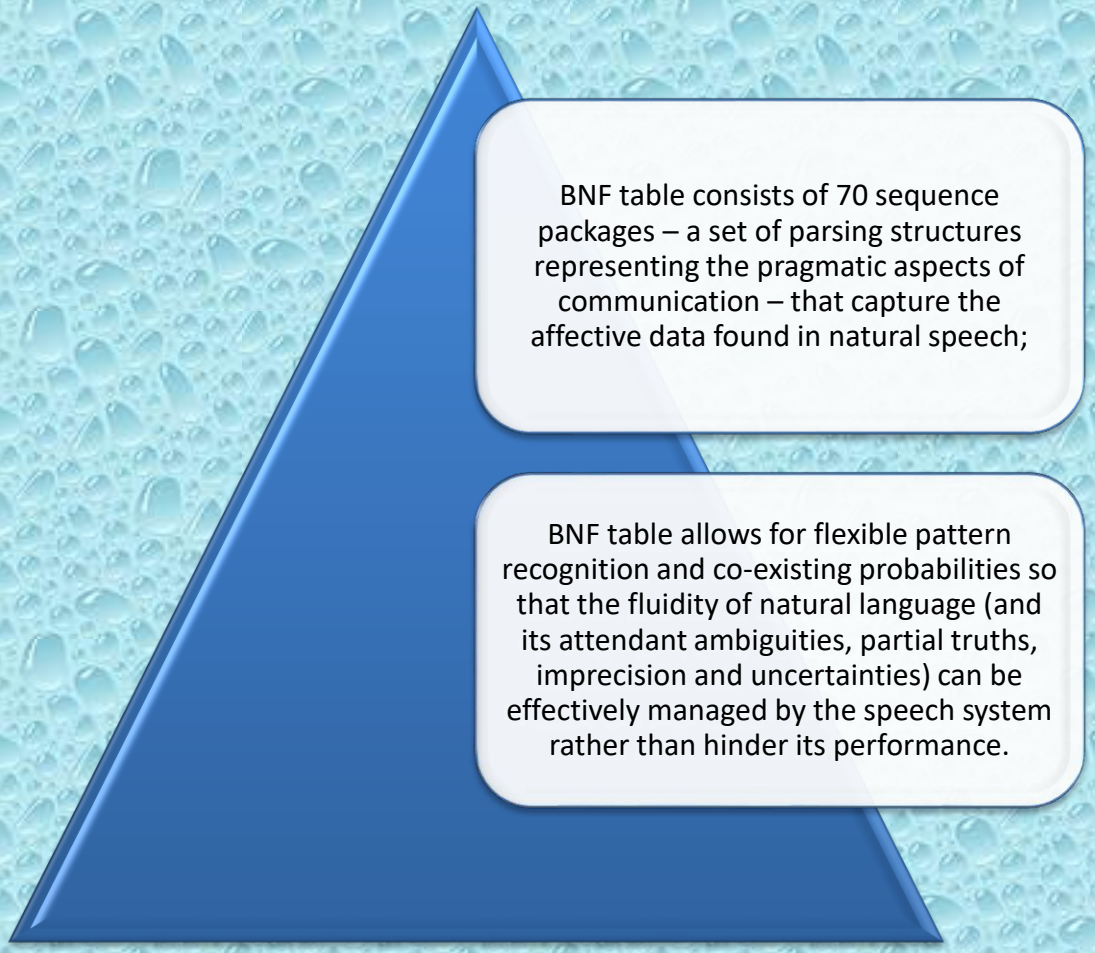


METHODOLOGY

Sequence Package Analysis (SPA) is built on a set of parsing structures – consisting of non terminals: *context-free grammatical units and prosodic features that capture the pragmatic aspects of communication: intent of speaker, status of speaker, inferences, context, interpretations, and connotations*

For the non terminals there is a corresponding list of *interchangeable* terminals: words, phrases, whole utterance(s).

Backus-Naur Form (BNF) Table of Sequence Packages



BNF table consists of 70 sequence packages – a set of parsing structures representing the pragmatic aspects of communication – that capture the affective data found in natural speech;

BNF table allows for flexible pattern recognition and co-existing probabilities so that the fluidity of natural language (and its attendant ambiguities, partial truths, imprecision and uncertainties) can be effectively managed by the speech system rather than hinder its performance.

SPA-DESIGNED PART OF SPEECH TAGGING: MACHINE LEARNING OF PRAGMATICS FROM SPECIAL ARRANGEMENT OF SYNTACTIC UNITS

To implement SPA, individual grammatical units are tagged with a POS tagger. For this stage an open source tagger (fasttagv2 by Mark Watson) was used with a changed lexicon that included the POS tags for SPA. We refer to this as an SPA unit.

Using sequence package analysis, we know that a specific order (arrangement, frequency and placement) of these units comprise an SPA component, which is built up from smaller, elemental grammatical units, referred to as SPA units.

The SPA units/components allow the system to check for caller frustration or suspicious behavior of terror suspects because SPA tagging captures the pragmatic aspects of dialog: speaker intent, inferences, and other aspects of context.

Govinda Keshavdas, Dep't of Electrical Engineering, University of Florida (Linguistic Technology Systems, Summer Intern , 2010)

IMPLEMENTATION

We have implemented a limited lexicon for SPA components by changing the existing lexicon of fasttag.

Some issues that we are working on are :

How large a lexicon is needed to achieve accurate results	How to make a robust system that gives accounts for all variations in conversational patterns, across different subject domains	Govinda Keshavdas, Dep't of Electrical Engineering, University of Florida

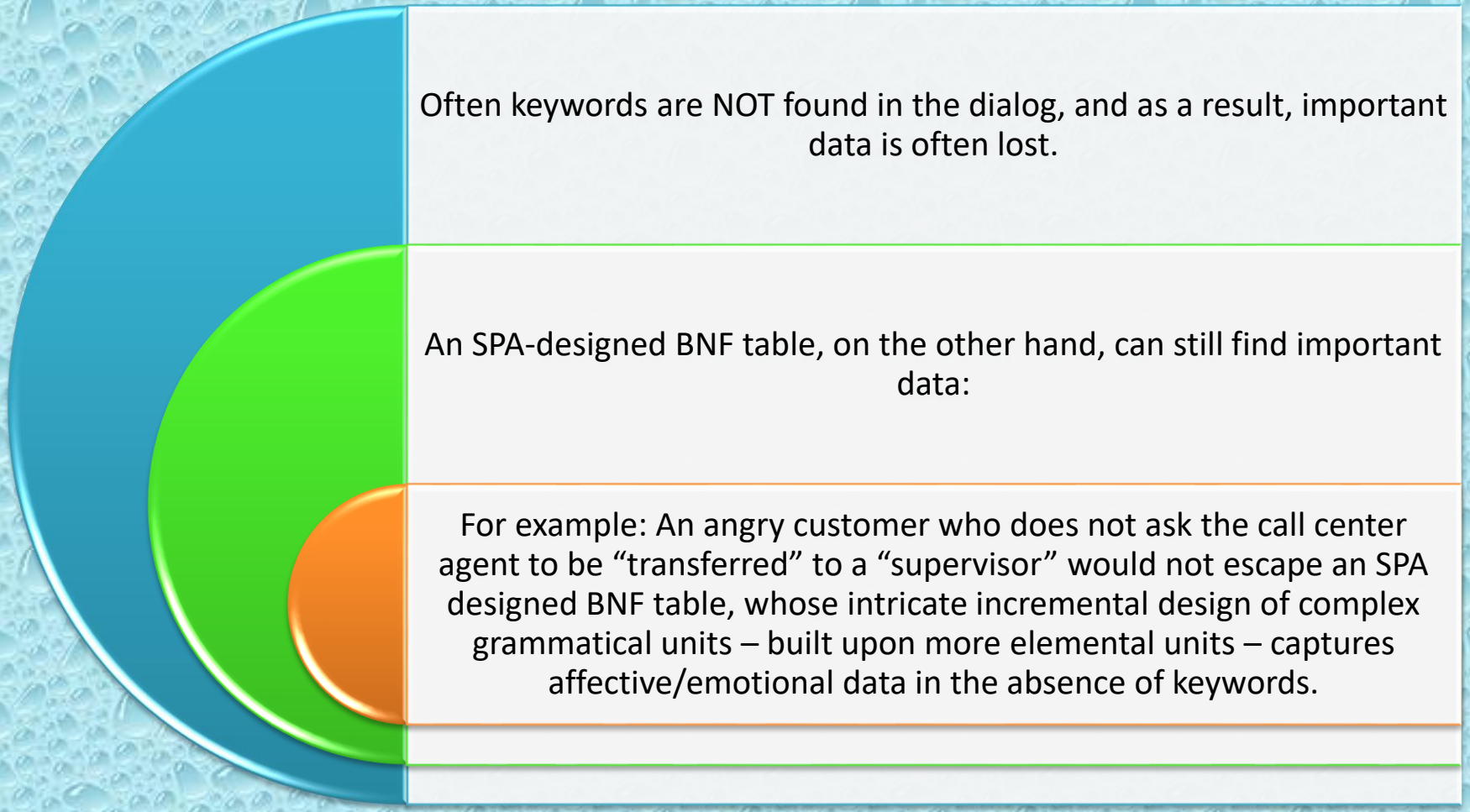
TESTING THE SYSTEM

INPUT TO SYSTEM: Transcript of recording which includes information about multi-channeling (which speaker is speaking at any given point in the dialog) pauses, intonation and other prosodic features.

PROCESS: INPUT->SPA units -> SPA components.

Govinda Keshavdas, Dep't of Electrical Engineering, University of Florida (Linguistic Technology Systems, Summer Intern 2010)

UTITLITY OF BNF TABLE OF SEQUENCE PACKAGES



Often keywords are NOT found in the dialog, and as a result, important data is often lost.

An SPA-designed BNF table, on the other hand, can still find important data:

For example: An angry customer who does not ask the call center agent to be “transferred” to a “supervisor” would not escape an SPA designed BNF table, whose intricate incremental design of complex grammatical units – built upon more elemental units – captures affective/emotional data in the absence of keywords.

PARSING FEATURES OF ANGRY CALLER DIALOG

A very angry complaint that does NOT contain keywords (e.g., “transfer” to “supervisor,” or “closing my account,” or “taking my business to competitor ‘X’”)

can be identified nevertheless by the natural accretion of the more elemental parsing features found in the dialog that match the units found in the pragmatically-based POS (part of speech) tagging program:

Assertions
Exaggerations
Declarations

The Human Mind as Role Model for SPA as a Soft Computing Method

1) Study How humans Resolve Natural Language Ambiguity and Imprecision present in Fluid and Dynamic Dialog

2) Design Simulacra to Resemble the Formulae that Humans Regularly Invoke to Understand Natural Language Dialog that is:

changeable, repetitious, elliptical – punctuated by deixes, anaphora and cataphora and other forms of linguistic expression that are only made clear by the dialog context

Pragmatic Competence

Pragmatic Competence: Speakers overcome ambiguity and imprecision of natural language dialog by taking into account the context of the utterance, knowledge about the status of those involved, and the inferred intent of the speaker, among other factors.

SPA is built on a set of pragmatically-based parsing structures – which provides speech systems with *pragmatic competence* that emulate the human's ability to understand the intended meaning of the other speaker, and other aspects of context.

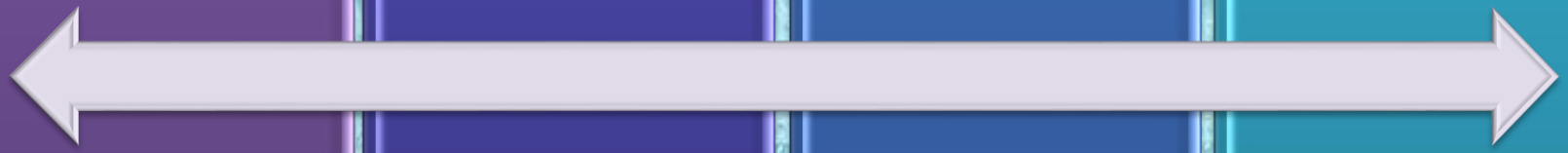
INDEXICALITY

Indexicality: Speakers work *actively* to find meaning (which is not fixed as in a dictionary) for the word or phrase that is inherently unique because it is embedded in context;

Speakers overcome indexicality – the uniqueness of each contextually embedded linguistic expression – by treating *new* material as an instance of a presupposed underlying pattern (shaped by culturally shared knowledge, ontology, contextual meaning, etc.)

SPA emulates this human process by mapping out the orderly sequences that form the context of the unfolding dialog:

the grammar the chart parser operates on has alternative patterns against which the new speech input can be matched so that the correct meaning will be assigned to the speech input.



Industrial and Environmental Applications for SPA

I. Commercial Sector:

- Mining recorded conversations of customers and call center agents to learn business intelligence and detect threats to customer retention
- Mining recordings of doctor/patient dialog to uncover important medical history data, lost in the roundabout way patients describe their symptoms

II. Public Safety:

- Mining Government Wiretap Recordings of Terror Suspects for Critical Intelligence Data
- Performing forensic speaker recognition of terror suspects and other criminals

Call Centers: Example 1

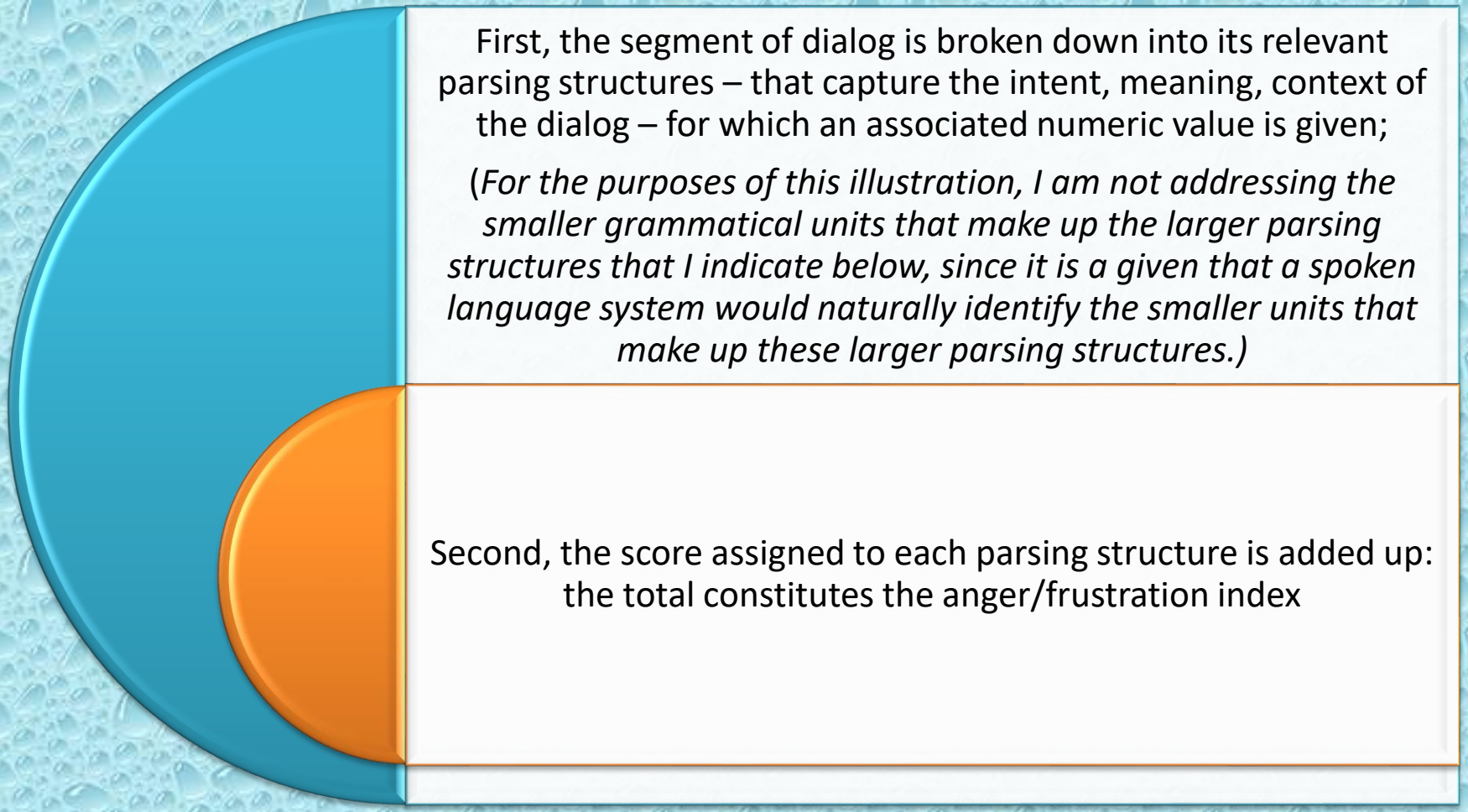
High Anger Level

(Punctuation symbols below are acoustic and not grammatical: question marks appear mid-sentence to indicate an upward query at that location point in the dialog; if inflection has risen an exclamatory marker is used)

- Caller: Absolutely unbelievable! What is your? name
- Agent: Mr. Smith
- Caller: Well! I intend to take this much further…This is just absolutely ridiculous!

Descriptors ("absolutely unbelievable" "absolutely ridiculous") have "high salience value" (they co-occur with the emotion class "anger" or "surprise" as opposed to a low salience value ascribed to more neutral words, such as "continue" or "yes"); yet there are still no "catch" phrases/keywords in dialog to signify an irate caller

How is Anger/Frustration Index Determined?



First, the segment of dialog is broken down into its relevant parsing structures – that capture the intent, meaning, context of the dialog – for which an associated numeric value is given;

(*For the purposes of this illustration, I am not addressing the smaller grammatical units that make up the larger parsing structures that I indicate below, since it is a given that a spoken language system would naturally identify the smaller units that make up these larger parsing structures.)*

Second, the score assigned to each parsing structure is added up: the total constitutes the anger/frustration index

COMPUTING TOTAL SCORE FOR ANGER/FRUSTRATION in Example 1

- Absolutely Unbelievable! <Exaggerative Qualifier> (8)
- What is your? name <Identification Request> (non sequitur; accusatory tone as indicated by displaced (mid- sentence) inflection) (9)
- Well! <Exclamation with Prosody> (7)
- I intend to take this much further…<Declarative Assertion> (9)
- This is absolutely ridiculous! <Exaggerative Qualifier> (8)
- Total Score for Customer Anger/Frustration Index: 41

Call Centers: Example 2

Moderate Anger Level

(the elliptical dots in the dialog below indicate there is no natural dropping of inflection at end of sentence but a continuation to the next sentence; the absence of a question mark at end of utterance indicates lack of upward inflection)

Caller: I'm just asking a question…I am just wondering whether or not I should install MS Word

As in the prior example, the caller does not use catch phrases or keywords to signify anger/frustration.

Computing Total Score for Anger/Frustration in Example 2

I'm just asking a question <Formulation> (5)

I'm just wondering <Repeat Formulation> (7)

Whether or not I should install MS Word <Question> (6)

Total Score for Customer Anger/Frustration Index: 18

While the anger/frustration index in this instant case is less than half the score of the prior dialog example, the speaker's use of two consecutive formulations – grammatical devices that permit a speaker to use some part of the dialog to "formulate" or "sum up" the unfolding activity (viz., the caller's asking of questions of a help-line desk agent) –*clearly indicate anger/frustration*

ADJUSTING TO THE FLUIDITY OF DIALOG AS SHOWN BY THIS EXAMPLE

The occurrence of two consecutive formulations as prefatory to a question signal a problem in the dialog:

A caller would not ordinarily preface his/her inquiry with “I’m just asking a question, I’m just wondering whether or not” – prefaces that appear more like a declaration than a simple request for help – unless he feels that his inquiry has not been properly addressed by the call center agent in the first place.

The second formulation is given a somewhat higher anger/frustration index than the first, as it indicates escalation in the speaker’s emotional state.

FLEXIBLE PATTERN RECOGNITION AND CO-EXISTING PROBABILITIES IN PRAGMATIC POS TAGGING

Had the question appeared in the dialog as a straightforward question (*"I don't know whether or not I should install MS Word?"*) without a prefatory set of formulations, the question would have been assigned the value of "1" – the lowest level on the anger/frustration index

But since the question followed two prefatory formulations it was assigned a moderate (to high) level of anger

This example shows how frequency and placement of parsing structures that comprise the sequence package can transform a simple inquiry into a (moderately) angry complaint, which is why it imperative for NLU (natural language understanding) algorithms to be guided by probabilities – keeping all of them simultaneously active at all times.

WIRETAPPED COMMUNICATIONS: CALCULATING COLLUSION INDEX

Speaker “A” is trying to educate Speaker “B” about a very important new meeting place right at the tip of the Brooklyn Bridge. Any confusion or misunderstanding about this meeting place could spoil the plans.

But Speaker “A” is very clever:

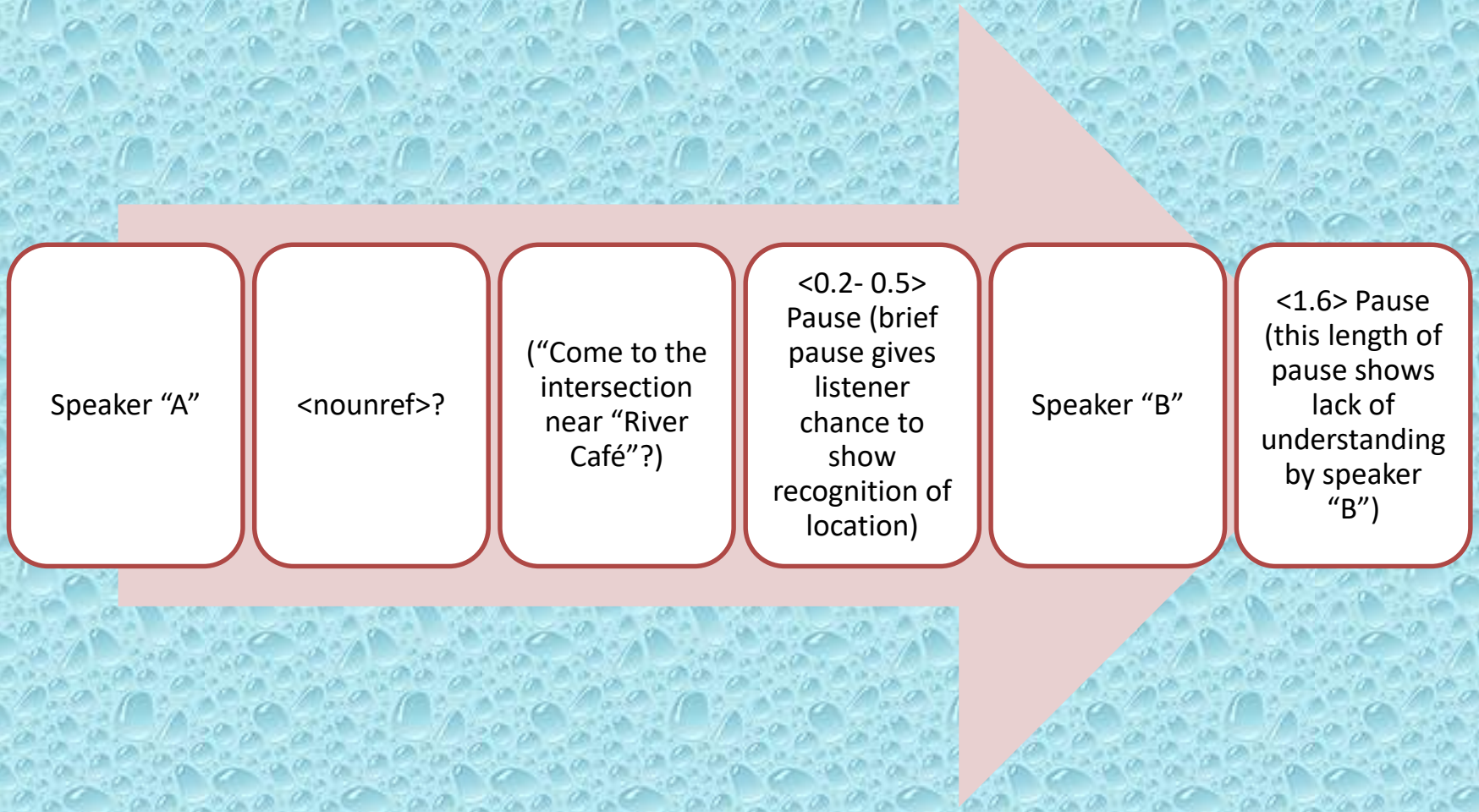
First, he stays away from buzz words (such as naming a bridge, a tunnel or a street).

Second, he refrains from using any prefaces or explanations to the other speaker about how vital it is to have these instructions - which are critical to carrying out the plans and operations - perfectly understood by the other speaker.

DIALOG SAMPLE

- Speaker A: Come to the intersection near River Cafe? upward intonation 0.2 - 0.5 pause
- Speaker “B”: 1.6 second pause
- Speaker “A”: You know (0.3-0.6) the thoroughfare with the big traffic light
- Speaker “B”: River Cafe, Yeah (!)

FINDING PARSING STRUCTURES IN DIALOG SAMPLE THAT POINT TO POSSIBLE COLLUSION



USE OF SPA POS TAGGING TO DETERMINE COLLUSION

Speaker “A”<clarif>(<clarif-pref><repaired-descrip>+)

<clarif-pref>“You know”

<repaired-descrip>+(<equal match><more specific>)“the thoroughfare”

intersection=thoroughfare: *same* level of generality in SPA lexicon

“with the big traffic light” – constitutes a *more specific* descriptor (*granularity* -- higher level of resolution or scale) to *match* “River Café” -- the source of the recognition trouble

PROOF OF COLLUSION FOUND IN SEQUENCE PACKAGE DATA

Speaker “B” <repeat-nounref> “River Café” - the source of the recognition trouble

followed by:

<*recognition-marker*> “Yeah”!

ANALYSIS

<repeat-nounref>+<recognition-marker>

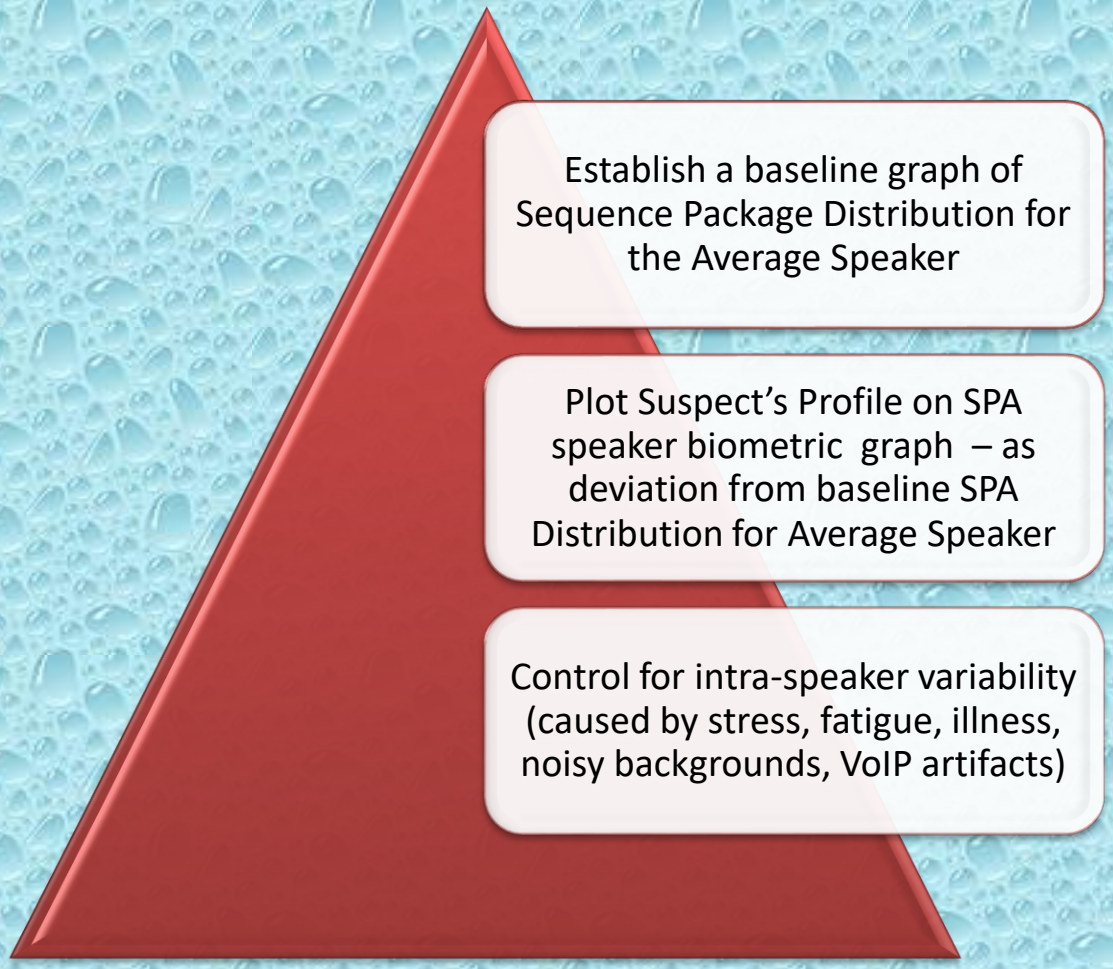
“River Café” “Yeah”

Had Speaker “B” simply produced a recognition marker (“yeah”) without mentioning the source of the trouble (“River Cafe”), there would be no clear indication to the other speaker that Speaker “B” now recognizes the importance of the meeting place.

This example shows how pragmatic features of dialog – the display of understanding of one speaker’s intent to another – can be mapped into parsing structures and built into an SPA POS tagging program.

COMPUTING COLLUSION INDEX SCORE

HOW TO ADD SPA TO SPEAKER BIOMETRICS

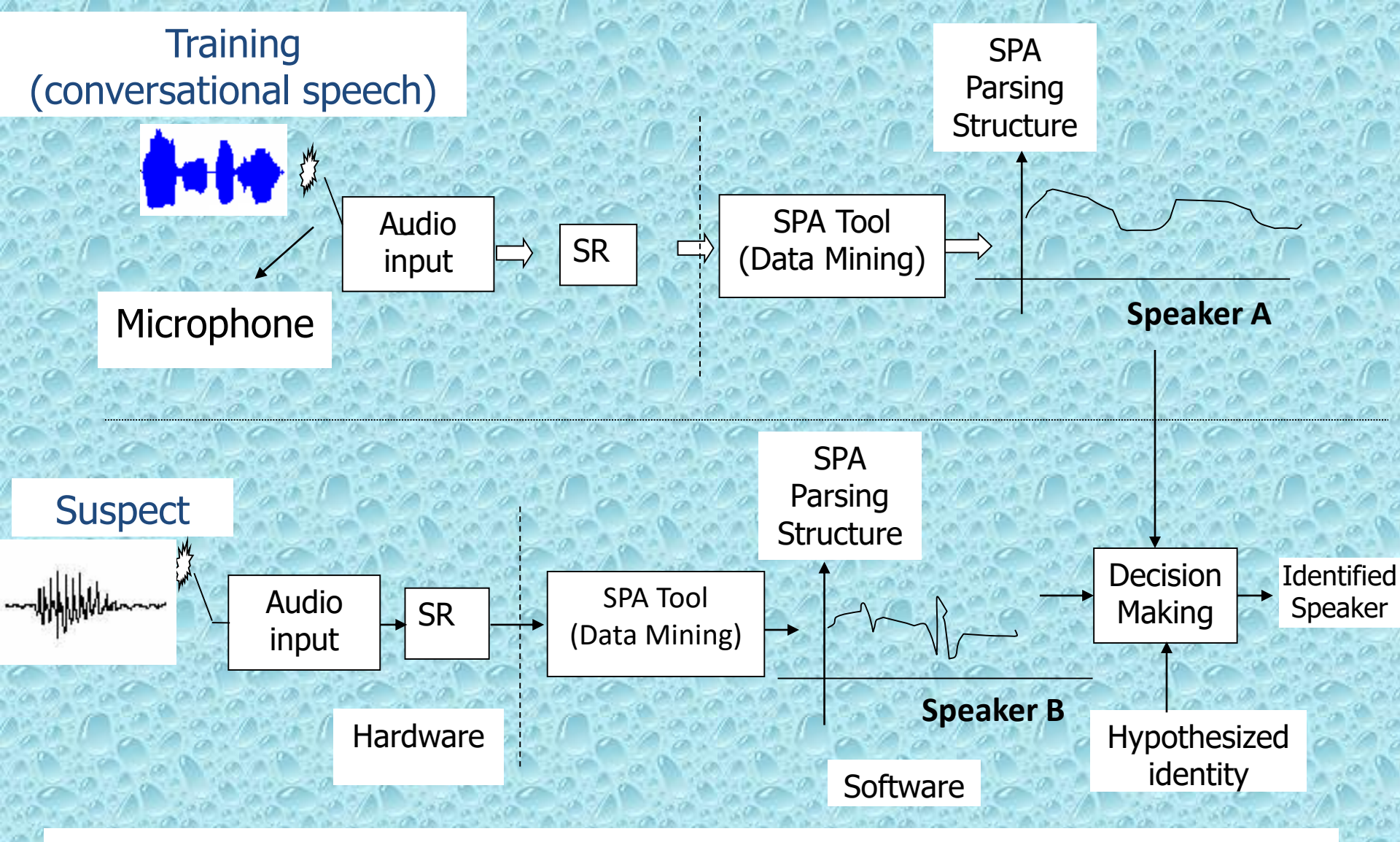


SEQUENCE PACKAGE PATTERNS AS A FEATURE OF VOICE RECOGNITION

Suspect's voice has normal formant (acoustic resonance), average nasality, prosody, etc.; as a result suspect can be easily mistaken for another speaker

On SPA graph, suspect shows deviation from baseline SPA distribution:

- Marked increase in pre-request/apologetic terms ("I'm sorry to bother you…"; "I was just wondering if…"; "May I ask you a question?") above average range
- Marked increase in topical digressions, as evidenced by use of topic displacement markers ("By the way,")



Speaker identification and verification

Professor Hemant Patil (Assistant Prof. DA-IICT)