Justifying Corpus-Based Choices in Referring Expression Generation

Helmut Horacek

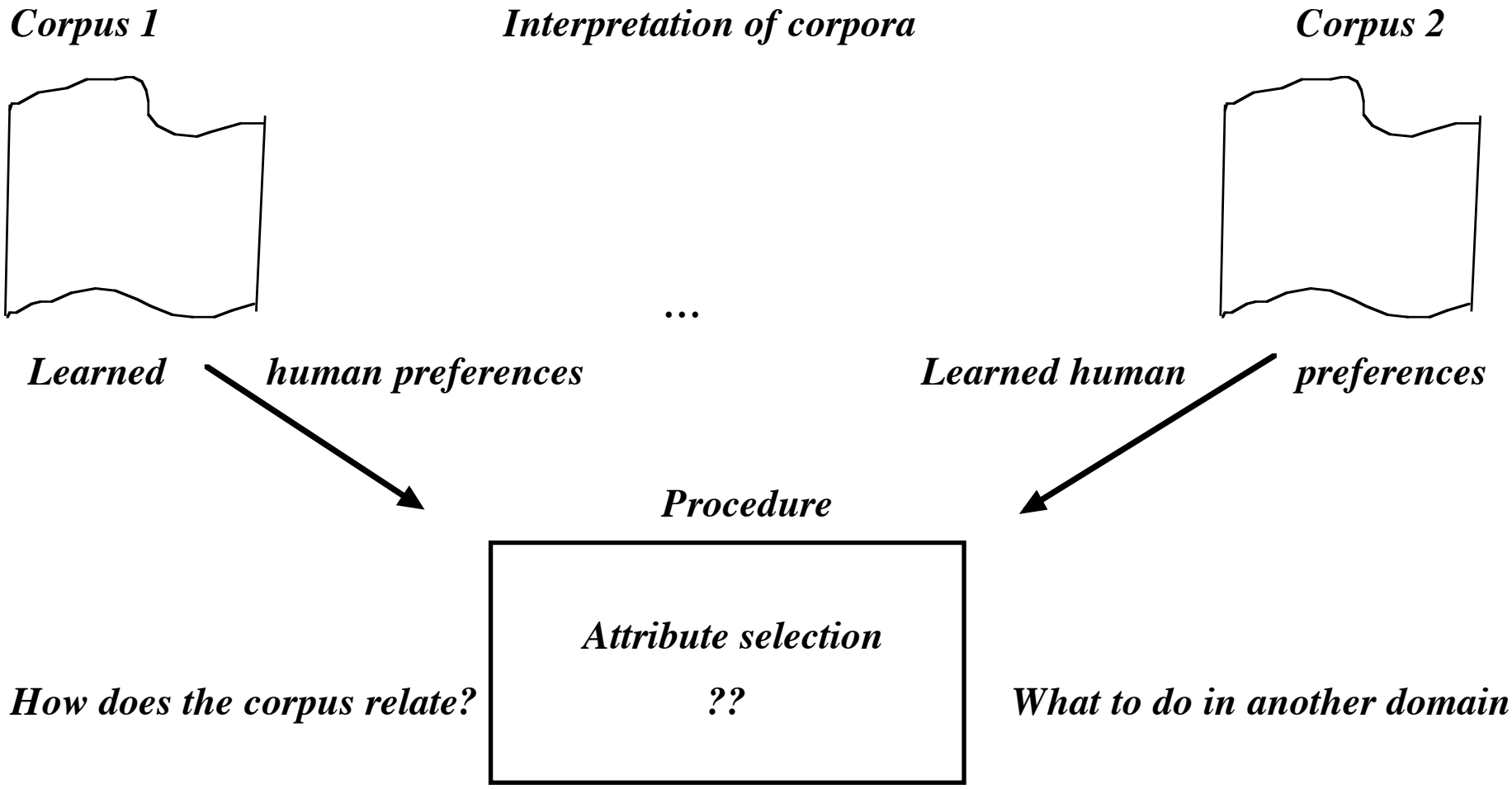
DFKI GmbH &

Saarland University

Recent Advances in Natural Language Processing (RANLP) - 2013

Hissar, Bulgaria, 7.-13.9.2013

CHOICES IN REFERRING EXPRESSION GENERATION



MOTIVATION

Approaches in the 1990s

Emphasis on algorithmic/computational issues

Instantiation and use of essential control parameters left unspecified

Approaches from 2000 on

Corpora generated through controled experiments

Selection procedures based on learning algorithms that interpret a partial corpus

Excellent results, but restricted to domain/specificities of the corpus addressed

Goal

Adopting empirical insights about human proferences

Abstracting from idiosyncracies of individual corpora

Expressing preferences in task oriented categories,

widely domain-independent, cognitive meaningful terms

CONTENT

PREVIOUS WORK

Algorithmic perspective

Settled in favor of the incremental algorithm [Dale and Reiter 1995]

Building corpora to mimick human preferences by learning procedures

- **TUNA corpus [van Deemter et al. 2012],**

  intensibely used for challenges [Gatt and Belz 2010]

- **GRE3D3 and GRE3D7 [Viethen and Dale 2008, 2011] - situationanl 3D**

Examinations of principles underlying attribute selection choices

- **Expressed in terms of rules [Jordan and Walker 2005]**
- **Role of visual context [Viethen et al. 2010]**
- **Adapting previous references rather than constructing a new expression [Viethen, Dale and Guhe 2011]**

THE BASIC IDEA

Method

Expressing relations between situations and expressions chosen

Established between properties of situations and compenents of expressions

Aggregated over similar situations

Assumptions

People's choices can be characerized in terms of components of expressions

Choices depend on *salience* of properties *and* on their contribution to *identification*

Novelty

The task of *identification* plays a *crucial role*

Simple or difficult cases, good or bad contribution of salient properties

CATEGORIES OF ATTRIBUTES

Obligatory elements

Attributes that must be chosen in some sort of situation (human almost always do)

Exclusive alternatives

Two attributes where one of them but not the other must be chosen in some sort of situation (human almost always choise exactly one of them)

Optional elements

Attributes that may be chosen in some sort of situation (human sometimes do)

Contextual factors

Leading to preferences in choosing among exclusive alternatives or distinguishing situations from others where optional elements are chosen or not

***Test through* aggregation over sets of situations (similar in the indentification task)**

APPLICATION – CORPUS USED

TUNA corpus

Organization

3x5 square grid

1 intended referent, approximately four distractors

Domains

Furniture

category, color, size, orientation

Scientists

beardedness, wearing glasses, age, hair(color), tie, …

CATEGORIES TESTED

Subcategorization of attributes

1. the type

2. most salient attributes (color, beardedness, wearing glasses)

3. location

4. remaining attributes

Contribution to identification – an attribute

1. allows the identification by itself

2. does it together with the type attribute

3. does it together with the type attribute and a most salient attribute

4. neither of these

RESULTS

Situations	***Regularity***
furniture domain	**obligatory (color)**
distinguishing (type+color+orientation)	**alternatives (position, orientation)**
distinguishing (type+color+size)	**alternatives (position, size)**
applicable (beardedness)	**optional (beardedness)**
distinguishing (beardedness)	**obligatory (beardedness)**
distinguishing (hair color)	**alternatives (hair color, position)**

DISCUSSION

Granularity and precision – rules are

More concrete than principles tested on the basis of controled experiments

Less specific than results obtained by learning methods

Domain-dependency

The role of identification contribution and relation between salient attribute:

***color* obligatory in entire furniture domain (as opposed to *hair color*)**

the role of *beardedness* and *wearning glasses* depend on role in identification

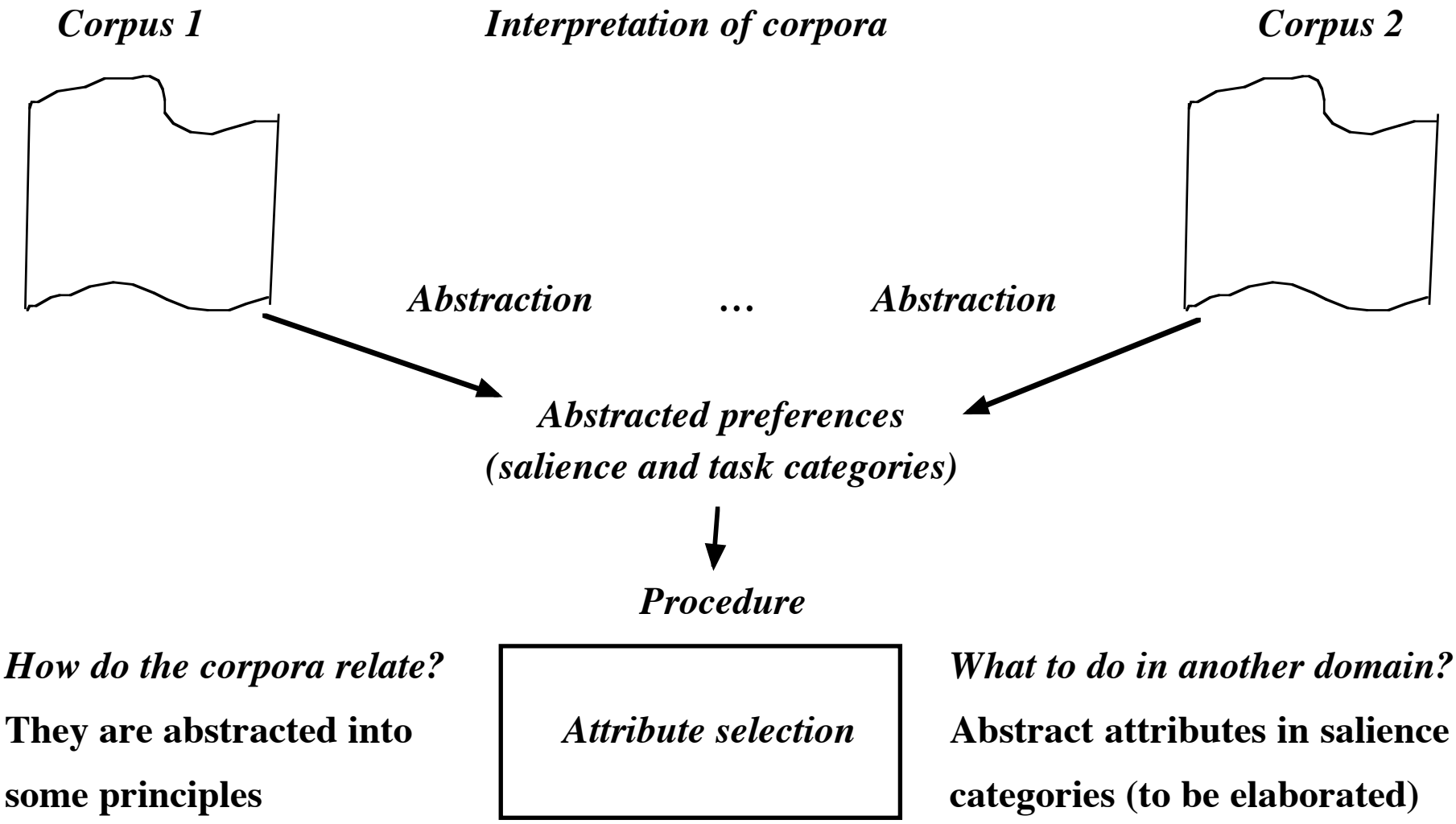
Towards transferring to other domains – assumptions

People behave similarly in similar situations (easy or difficult identification task)

People behave similarly in comparable perception circumstances (attribute salience)

Salience can be reasonably generalized across situations and domains

THE VISION


Corpus 1
Interpretation of corpora
Corpus 2
Abstraction
…
Abstraction
Abstracted preferences
(salience and task categories)
Procedure
How do the corpora relate?
They are abstracted into some principles
Attribute selection
What to do in another domain?
Abstract attributes in salience categories (to be elaborated)

CONCLUSION

Approach

Finding out relations between task-relevant properties and attributes chosen

Application to the TUNA corpus

Results

Some principled rules extracted from the corpus

Some discrepancies regarding domains and the role of salient attributes found

Extensions

Application to other corpora

Towards automation of choosing and testing aggregations of situation

More fine-grained descriptions of regularities