

# **Extraction and Formalization of Support Verb Paraphrases from Corpora**

**Applications in Machine Translation**

**Anabela Barreiro**

anabela.barreiro@nyu.edu

**Universidade do Porto**

**New York University**

## **Advisors**

Belinda Maia, FLUP/CLUP/Linguatca  
Adam Meyers, New York University



# Outline

- Aim of the Study, Goals and Best Scenario
- Motivation
  - Why Bilingual Paraphrases?
  - Why Support Verbs?
  - Importance of Multi-word Identification for Machine Translation
- Background
  - Task Definition
  - The Corpus
  - Overview on Machine Translation
  - Overview on Paraphrasing
- Methodology
  - Starting Point
- Conclusions and Prospectives

# Aim of the Study

This study performs linguistic analysis and computational formalization of constructions where **bilingual short paraphrases** occur, with the main purpose of using them in MT

Examples:

*make a call = fazer um telefonema = fazer uma chamada = telefonar*  
*make a living of N = governar a vida de N = viver de N*

*fazer uma investigação = make an investigation = perform an investigation = investigate*  
*fazer uma visita = pay a visit = visit*

# Plan by Steps

1. assemble a bilingual corpus of lexicalizations “verb + complement” where the main verb has a weak semantic value, the so called “support verbs”, or “light verbs”, such as “*take*” in “*take a walk*” (EN) or “*fazer*” in “*fazer uma apresentação*” (PT)
2. identify and build EN-PT and PT-EN syntactic-semantic correspondence rules for **support verb constructions** (SVCs) and study transfer mechanisms between the two languages
3. create rules and requirements to find short paraphrases

# General Goals

- Create sharable bilingual lexical resources (compile sets of SVCs translations and equivalent paraphrases for a pair of languages)
- Build a tool for automated extraction of short paraphrases
- Make methodology be suitable for anyone working on MT or bilingual lexicography
- Apply results to an MT system

# Value of the Results

- Machine Translation
  - valuable resource for PT-EN or EN-PT MT system translation pairs
  - high significance for similar work on other languages or language pairs
- Paraphrase-lexica represent a valuable resource for many other NLP-applications
  - Information Extraction (IE)
  - Text Mining (TM)
  - Information Retrieval (IR)
  - Summarization (S)
  - Question-Answering (Q&A)

# Value of the Results

## Importance of Paraphrases in Related Fields

- Literature
- Translation
- Linguistics
- NLP

# Why Paraphrases?

## Literature

- Closely related to analysis, interpretation and synthesis
- Recognized as repeating content but conveying it with different expressions

## Translation

- Translation theories are deeply concerned with and involved in the field of paraphrasal existence and use as well. In general terms, free translation is a set of paraphrases which represent the same idea conveyed from a source language into a target language



# Why Paraphrases?

## Linguistics

- Central to many linguistic theories which consider it both an integral part of a speaker's linguistic competence and a central tool for language modeling
- Associated with synonymy. At the lexical level, synonymy concerns words and phrases. At the syntactical level, synonymy means keeping the same meaning in a sentence with a different structure

Barzilay & McKeown, 2001 pointed out: “Many linguists (Halliday, 1985; de Beaugrande and Dressler, 1981) agree that paraphrases retain “approximate conceptual equivalence”, and are not limited only to synonymy relations.”

# Linguistic Theories

- Meaning-Text Theory (MTT) (cf. Mel'čuk, 1988)
  - views language as a “translation system”, a series of modules through which a meaning is converted into a well-formed sentence
  - Paraphrasing rules ensure that all the sentences one can produce to express a particular meaning are semantic equivalents of one another

Milićević, 2005

- proposes new types of paraphrasing rules capable of handling paraphrase-semantic equivalence rules, operating at the semantic level of utterances representation
- Generative Transformational Grammar (cf. Chomsky, 1957, 1965)
    - transformations are instances of paraphrasal representation where paraphrases are surface structures for the same deep structure
    - Ex: transformational rules such as active to passive voice

# Why Paraphrases?

## NLP

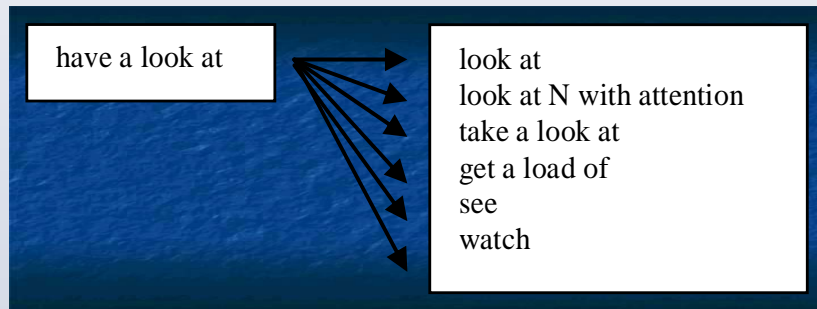
- As initially stated, the study of paraphrases also plays a very important role in natural language interpretation and generation and is used in many different NLP applications
- Some of today's NLP systems use automated methods to collect paraphrases

# NLP Relevant Work

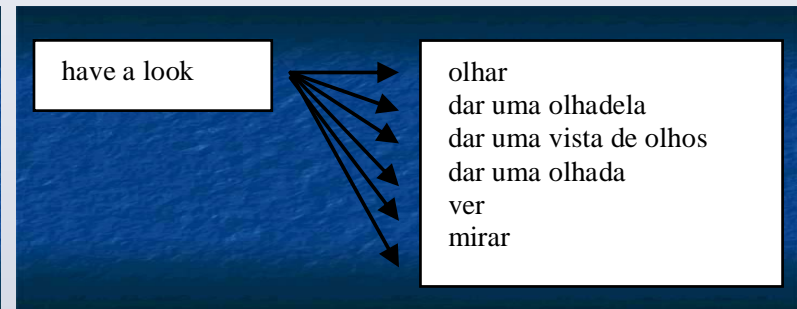
- Regina Barzilay's thesis (2003)
- Proceedings of the ACL 2003 Workshop on Paraphrasing
- Other thesis
  - Mark Dras (1999)
  - Florence Duclaye (2003)
- Papers
  - Barzilay & McKeown, , 2001
  - Dolan et al., 2004,
  - Poibeau, 2004
  - Pang et al., 2003
  - Barzilay & Lee, 2003
  - Shinyama, 2003
  - Quirk et al., 2004
  - Bannard & Callison-Burch, 2005
  - Resnik, 2005
  - Satoshi, 2005
- People
  - Dekang Lin, Lillian Lee, Kevin Knight, Satoshi Sekine, Yusuke Shinyama, Hua Wu

# Paraphrase (within the scope of this work)

- substitute technique to express the same meaning in a phrase
- replacement by semantic-equivalent phrase or word



Monolingual



Bilingual

# Why Support Verbs?

SVCs present translation difficulties for several reasons:

- 1- SVerbs in EN and PT are of different type (eg. *fazer adiamentos* => *to give extensions*; *fazer perguntas* => *to ask questions*; *fazer o exame* => *to take the exam*; *fazer uma visita* => *to pay a visit*). Even though some SVerbs can be translated literally (eg. *fazer dinheiro* => *to make money*), most of them have no literal translation (eg. *fazer uma operação* => *to have an operation*; *fazer perguntas* => *to ask questions*)
- 2- some SVCs are frozen in one or in both languages, being considered fully idiomatic (collocations) (eg. *fazer boa figura* => *to do well*; *fazer o sangue subir à cabeça* => *make the blood rush to one's head*; and less common *fazer o papel de pai* => *to play the father*)
- 3- several expressions in PT either require SVerbs or the SVC is more often used and/or preferred, simply corresponding in EN to a lexical verb (eg. (?preferred) *fazer publicidade* = *publicitar* => *to advertise*; (?preferred) *fazer um esboço* = *esboçar* => *to draft*; *fazer compras* => *to shop*; (?preferred) *fazer a barba* = *barbear* => *to shave*)
- 4- many SVCs match with lexical verb variations, expressing exactly the same meaning (eg. *tirar fotocópias* = *fotocopiar* => *to make copies* = *to xerox* = *to photocopy*; *fazer uma doação* = *doar* => *to make a donation/donations* = *to donate*).

# Why Support Verbs?

- Support verbs are interesting to our research because they often represent phrasal alternatives = paraphrases

Simple example

*to pay a visit = to visit*

- The importance of studying these constructions is based on their abundance in language and their study is essential for MT
- An MT system cannot ignore the the meaning of “*pay*” in “*pay attention*”, for instance.
- We assume the fundamental principle that elementary sentences/phrases and not the words carry units of meaning (meaning is not compositional). We also assume the relation of nominalization as defined by Harris, 1964

# Relevant Work on SVCs/Light Verbs

- The problem of the translation of SVCs is known and being tackled since the sixties
- Gross, 1981, Giry-Schneider, 1987 (French school of linguistics)
- Dras, 1995 (Anglo-Saxon school)
- Grimshaw & Mester (1988), Larson (1988), Hale & Keyser, 1991, 1993, Chomsky, 1995, 1998, 1999
- Mel'čuk, 2003 (paraphrasing widely published in French)
  - Mel'čuk notion of "lexical function" (= formalism for description and use of combinatorial properties of individual lexemes) allows for an elegant solution of the problem by shifting the work load to the compilation of monolingual collocation dictionaries



# Relevant Work on SVCs/Light Verbs

- Lexical functions are used, e.g., in the wide-coverage MT-system ETAP; see the numerous publications by Apresjan, Boguslavsky and their co-workers
- The Grenoble MT-group is also working with LFs across languages (e.g., the project Papillon)
- Thierry Fontanelle and the Decide project
- FrameNet
- Central references such as Allerton and Grimshaw
- Alonso Ramos (2004) in Spanish discusses in detail the representation of SVCs in the framework of Mel'čuk: *Las Construcciones con Verbo de Apoyo*, Visor Libros, Madrid.

# Importance of multi-word for MT

## State of the Art MT

- MT nowadays – are we there yet?
- Studies analyzing MT progress and comparative evaluations of systems from the 1970's with today's systems don't conclude that the new systems are significantly better than the old ones (see Hutchins, 2003)
- Where & why MT still fails and what we must DO  
See Challenges for MT (Barreiro & Ranchhod, 2004)

# Several MT outputs

English sentence:

*Then the operatives are more likely **to make mistakes** when set-ups are constantly changing.*

Manual translation from Compara:

*E os mecânicos estão mais sujeitos a **cometerem erros** quando os sistemas estão constantemente a mudar, e isso conduz a um aumento de desperdício.*

MT outputs:

Translation Engine	Result
FreeTranslation	Então o operativo são mais possível <b>cometer erros</b> quando organizações constantemente mudam.
Systran	Então os operatives são mais prováveis <b>fazer erros</b> quando as instalações estão mudando constantemente.
E-T Server	Então os operadores são mais prováveis para <b>tornar erros</b> quando as situações estão mudando constantemente.
Amikai	Então as operações são mais provável <b>cometer erros</b> quando jogos constantemente está mudando.

<http://www.linguateca.pt/METRA/>

(cf. Sarmiento et al., 2005, Sarmiento, 2005, and Maia & Barreiro, 2005 )

# Questions to Be Answered

1. How research into one linguistic phenomenon would be of interest to anyone but another linguist?

2. How would a study of paraphrases based on SVCs and their lexical equivalents benefit MT?

Exs: **take(vt) place = occur(vi) = happen(vi)**

“most of the action must **take place** in the same studio set” => “a maior parte da acção deve **desenrolar-se** no mesmo espaço cénico”

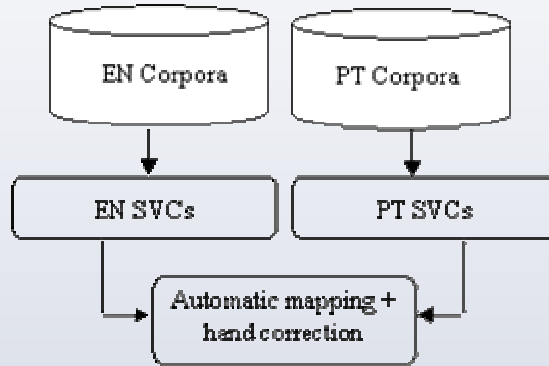
**make(vt) reparation = compensate(vi)**

“It was not too late **to make reparation** for that. => Não era tarde demais para **recuperar** o mal cometido.

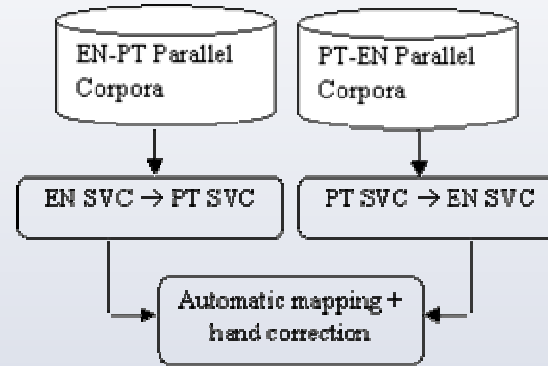
# Methodology

- Methods for identifying candidate constructions in corpora
- How are the paraphrases arrived at?
- How are these pairs constructed?

# Methods for SVC Extraction and Treatment



**Monolingual SVC extraction and treatment**



**Bilingual EN-PT and PT-EN extraction and treatment**

## Results of Application of a Simple Nooj SVC Grammar to a Monolingual Text

● concerned in it were	taking their pleasure/	quietly, and they were not of the sex which is supposed to
● re over, and he was	taking the rest/	that precedes the great rest. He had a narrow, clean-shaven
● sion Ralph Touchett	took no interest/	whatever; he stood with his hands in his pockets, looking g
● in she prepared to	take her departure/	. "Your sister must be a great gossip. Is she accustomed to
● rgument; you always	take the opposite ground/	. I don't see what you've against her except that she's so o
● ; she has evidently	taken a great fancy/	to her." "What is it you wish her to do?" Edmund Ludlow as
● s well as resolute,	took his way back/	to his lodging with the feeling of a man defeated. He was n
● American school and	took a degree/	at an American university, after which, as he struck his fa
● tant, remounted and	took his departure/	before dinner; and an hour after this meal was ended Mr. an
● dea that she hadn't	taken the trouble/	to listen to it. Of course I talk like an American--I can't
● unted them. I never	took much notice/	of the classes. That's the advantage of being an American h
● new. I've heard you	take such opposite views/	." "I'm on the side of both. I guess I'm a little on the si
● pon her, and Isabel	took a fancy/	to the young ladies, who appeared to her to show a most ori
● hat centre that she	took the train/	for the station nearest to Gardencourt, where Isabel and Ra
● "So few people will	take the trouble/	." Miss Stackpole, in this undertaking, appeared to shrink
● I see. And then?" "	Take right hold/	of something." "Well, now, what sort of thing?" "Anything
● n!" CHAPTER 11 He	took a resolve/	after this not to misinterpret her words even when Miss Sta
● it. He was about to	take a step/	which would astonish all his friends and displease a great
● they stopped and he	took her hand/	. "Ah, Lord Warburton, how little you know me!" Isabel said
● it seems to me I'm	taking the best way/	. If you'll be my wife, then I shall know you, and when I te
● efinately yet; I've	taken a little time/	to think of it, because that seems more respectful. But I s
● ung man from Boston	take positive possession/	of her. The sentiment in which she sought refuge after read
● rsecute her he must	take the consequences/	; foremost among which was his being left to perceive how li
● d Ralph Touchett to	take a walk/	with her in the garden, and when he had assented with that
● girl she was. She's	taking different views/	, a different colour, and turning away from her old ideals.
● u don't know how to	take such ideas/	. Mr. Goodwood does, however; and that's what I like about h
● nize; but it so far	took the edge/	from too sharp a recognition that, in spite of his suspecti
● reality and he had	taken no pains/	to ask questions the answers to which would involve only th
● hat his imagination	took a flight/	and that he remained in his chair a long time beyond the ho
● of course if you'll	take a companion/	--some decayed gentlewoman, with a darned cashmere and dyed
● say. If you do, you	take a great responsibility/	." "You frighten me a little, but I think I'm right," said
● for being quick to	take a hint/	that was meant as good advice. His advice had perhaps helpe
● n waves, her glance	took a backward flight/	. It rested upon two figures which, in spite of increasing d
● enquired. "Yes, we	take the train/	again. We've so much to do la-bas." "Are you not tired?"
● due to any one who	took the tone/	of authority; and she was a passive spectator of the operat
● surance the visitor	took his leave/	; after which Isabel fully expected her friend would scold h
● rle. Ralph Touchett	took his refreshment/	wherever he could find it, and he would not have forgotten
● ther well enough to	take their ease/	; but every now and then Isabel heard the Countess, at somet
● w, the golden light	took a deeper tone/	, and on the mountains and the plain that stretched beneath
● no talents even; I	took my measure/	early in life. I was simply the most fastidious young gentl
● Madame Merle's eyes	took the same direction/	; then she looked serenely at her neighbour. "You know I nev

# EN-PT Parallel Corpora

- COMPARA

<http://www.linguateca.pt/COMPARA/>

(cf. Frankenberg-Garcia & Santos, 2003)

- corpus based on Portuguese-English and English-Portuguese source texts and translations (literature)
- over one million words for each language



## Results of your search

You are welcome to use the results of your COMPARA search queries for research and education, provided the source is acknowledged. To cite specific texts from the corpus, click on the blue code next to them for their full reference. To cite the corpus as a whole, please refer to: COMPARA 6.7.1 anotado <http://www.linguateca.pt/COMPARA/> [14-Feb-2006]

### Short description of the corpus used in the present search:

Portuguese words	English words	Alignment units
1347280	1451766	90801

23 instances found.

Results for the search  
 "took place"  
 parallel concordances

### Concordance

Search string: "took" "place".

<a href="#">EBDL2</a> (929):	Sutcliffe tapped the side of his long nose with his index finger as he uttered these words, dislodging his spectacles and causing them to crash into his tea-cup -- for this conversation <b>took place</b> in the Senior Common Room, not long after Robyn's arrival at Rummidge.	Sutcliffe esfregara de lado o seu comprido nariz com o dedo indicador ao proferir estas palavras, deslocando os óculos e fazendo-os cair na chávena de chá, pois esta conversa teve lugar na sala dos professores, pouco tempo depois da chegada de Robyn a Rummidge.
<a href="#">EBDL4</a> (1255):	It <b>took place</b> on a fine October day.	Realizou-se num belo dia de Outubro.
<a href="#">EBKI1</a> (1260):	A certain sequence of events <b>took place</b> earlier this evening, and as a consequence, when I last left Mr Brodsky a few hours ago, he was sipping a small glass of whisky.	Ao princípio da noite verificou-se uma certa sequência de acontecimentos e, como resultado disso, quando deixei o Sr. Brodsky, há poucas horas, ele estava a beber um pequeno <i>whisky</i> .
<a href="#">EBKI2</a> (307):	So much so that after his arrival at Darlington Hall, even the brief exchanges necessary to communicate information relating to work <b>took place</b> in an atmosphere of mutual embarrassment.	A tal ponto que, depois da sua ida para Darlington Hall, até as breves trocas de palavras necessárias para transmitir informações relacionadas com o trabalho se faziam numa atmosfera de embaraço mútuo.
<a href="#">ESNG3</a> (43):	It did not so much as occur to her that it could have been possible to talk to other women about what was asked in the conversations with her husband that never <b>took place</b> .	Nem sequer lhe ocorria que seria possível falar com as outras acerca do que perguntava nas inexistentes conversas com o marido.

EN: Sutcliffe tapped the side of his long nose with his index finger as he uttered these words, dislodging his spectacles and causing them to crash into his tea-cup - - for this conversation **took place** in the Senior Common Room, not long after Robyn's arrival at Rummidge.

PT: Sutcliffe esfregara de lado o seu comprido nariz com o dedo indicador ao proferir estas palavras, deslocando os óculos e fazendo-os cair na chávena de chá, pois esta conversa **teve lugar** na sala dos professores, pouco tempo depois da chegada de Robyn a Rummidge.

EN: It **took place** on a fine October day.

PT: **Realizou-se** num belo dia de Outubro.

EN: I don't know what **took place**.

PT: Não sei o que **se passou**.

Figure 4: COMPARA search results

# More EN-PT Parallel Corpora

- **Europarl**

<http://people.csail.mit.edu/koehn/publications/europarl/>

- **NATools**

<http://eremita.di.uminho.pt/albin/nat-search.cgi>

# NOMLEX Extension

1. Identification of the support verbs
2. Mapping support verb arguments to nominalization arguments
3. Categories of support verbs – semantic content
  - Special complements of support verbs plus nominalizations
  - Classification of support verbs
4. Dictionary build

# NOMLEX Extension

**Co-occurrence of nominalizations with support verbs:**

***"launch an attack"***

***"take a walk"***

- **Design extended nominalization entries to capture information about these support verbs (Proteus Project Memorandum)**
- **Annotate all the nominalizations in the Penn Tree Bank**
- **Extend and validate the entries in NOMLEX**

```

(NOM :ORTH "destruction"
  :PLURAL *NONE*
  :VERB "destroy"
  :NOM-TYPE ((VERB-NOM))
  :VERB-SUBJ ((N-N-MOD)
              (DET-POSS)
              (PP :PVAL ("by"))))
  :VERB-SUBC ((NOM-NP :SUBJECT ((N-N-MOD)
                                (DET-POSS)
                                (PP :PVAL ("by"))))
              :OBJECT ((DET-POSS)
                       (N-N-MOD)
                       (PP :PVAL ("of"))))
              :REQUIRED ((OBJECT :DET-POSS-ONLY T
                                :N-N-MOD-ONLY T))))
  :DONE T)

```

NOMLEX entry for *destruction*

# Some SVC Classes

## PT Nouns Semantic Groups Co-occurring with Support Verbs

- MOVEMENT

With DAR - caminhada, curva(s), passeio, salto, volta, cambalhotas, pulos, etc.

With FAZER - pirueta, desvio, entrada, saída, pausa, etc.

- PHYSICAL AGGRESSION

With DAR - estalo, bofetada, sova, surra, golpe, cornada, pancada, murros, etc.

- DISEASES AND SYMPTOMS

With TER - acne, anemia, diabetes, lesão, ataque, infarte, gripe, constipação, depressão, crise, diarreia, febre, enjojo, sufoco, vertigem, arrepio, etc.

With APANHAR - gripe, constipação, depressão, etc.

- ACTIONS RELATED WITH SENSES

With DAR - abraço, beijo, gemido, grito, lambidela, olhada, sopro, suspiro, arrote, etc.

- PSYCHOLOGICAL AND EMOTIONAL

With DAR - ânimo, alegria, tristeza, desgosto, vontade, gosto, pânico, pena, satisfação, susto, terror, vergonha, etc.

With TER - angústia, covardia, prejuízo, dúvida, fobia, inquietude, moléstia, nostalgia, preocupação, receio, suspeita, emoção, etc.

With FAZER - medo, impressão, tentativa(s), etc.

- POWER AND INFLUENCE

With DAR - aprovação, autorização, aviso, ajuda, benefício, conselho, consentimento, cumprimentos, exemplo, licença, ordem, orientação, etc.

- MORAL AND FEELINGS

With DAR - afecto, amor, amparo, carinho, compreensão, consolo, escândalo, graças, prêmio, presente, prenda, etc.

# Some SVC Classes

## PT Nouns Semantic Groups Co-occurring with Support Verbs

- **BODY CARE**

With TOMAR - banho, duche, etc.

With FAZER - penteado, maquilhagem, barba, etc.

With DAR - banho, etc. With

- **INFORMATION-OPINION**

With DAR - assessoria, pormenores, explicação, informação, interpretação, mostras, opinião, orientação, resposta, declarações, etc.

With FAZER - relato, perguntas, advertência, juízo, raciocínio, recomendação, reflexão, sugestão, valorização, votos, recomendação, reflexão, acordo, raciocínio, declarações, etc.

With TOMAR – decisão

With TER - ideia, impressão, etc.

- **GREETINGS, CONDOLENCES, FELICITATIONS**

With DAR - bons dias, boas noites, boas tardes, parabéns, pêsames, saudação, honras, etc.

- **HOSTING**

With DAR - albergue, alojamento, asilo, hospedagem, refúgio, abrigo, etc.

- **REACHING A GOAL**

With DAR - alcance, captura, caça, etc.

- **FAST ACTIONS**

With DAR - pincelada, lavagem, esfregadela, etc.

- **USAGES**

With DAR - colocação, destino, emprego, encargo, etc.

- **OTHERS**

With DAR - alimento; calor, frio; início, origem; forma; força; leitura; vida; oportunidade; preferência, prioridade; realce; aso, riso; solução, etc.



# Some Support Verb Classes

## Dividing SVerbs by Semantic Content

Category	Description	Example
• :INCEP	beginning of event/state	<b>began</b> the evacuation <b>commence</b> with the attack
• :CONTIN	continuation of event/state	<b>proceeded</b> with the evacuation <b>kept</b> influence on John
• :FIN	completion of event	<b>completed</b> the evacuation the attack <b>ceased</b>
• :AFFIRM	content is affirmed	<b>admit</b> an accusation <b>maintain</b> an accusation
• :DISAFFIRM	content is disaffirmed	<b>deny</b> an accusation <b>reject</b> an accusation
• :NON-FACTIVE	state/action may not be a fact	<b>promised</b> John a gift <b>plan</b> the murder
• :NON-FACTIVE-INDEF	non-factive unless definite	<b>intend</b> an attack (non-factive) <b>intend</b> the attack (factive)
• :TRANSPARENT	does not fit in other category	<b>make</b> an attack <b>carry out</b> an attack
• :REALIZED	goal is achieved	the attack <b>succeeded</b> <b>approve</b> an appeal
• :NOT-REALIZED	goal is not achieved	the appeal <b>failed</b> <b>deny</b> an appeal

# Preliminary Conclusions

- SVCs and light-verb constructions refer to very particular phenomena, cross-categorisation with idioms

## Verb-object combinations

- *take N/Pron/somebody for a ride = take advantage of N/Pron/somebody*
- *take a close look at = look carefully = (inspect) = (investigate)*
- *bring N/Pron/something to light = to reveal N/Pron/something*
- *hold Pron head high = be proud*

# Preliminary Conclusions

We propose a classification of paraphrases in several categories:

- Simple paraphrase  
*make a confession = confess;*
- Implied meaning paraphrase  
*make a good impression = impress (where “good” is implied in the paraphrase)*
- Paraphrase with remote dependency  
*make a barricade = barricade SELF in PL*
- Paraphrase with change in the POS of some elements, like adjective to adverb  
*make a good scene = film N ADV\_well*  
*make the right decision = decide ADV\_correctly*  
*have a satisfactory answer = answer ADV\_satisfactorily*
- Paraphrase with complex transformations, requiring a “new” word  
*make an important disclosure = disclose something important*

**Feedback is welcome!**

**Thank you!**

Anabela

:)