Explicit objects and polysemy:
Do dictionaries get it right?

Stella Markantonatou[1], Antonis Anastasopoulos[2] and Yanis Maistros[2]

[1]Institute for Language and Speech Processing / Athena RIC, Athens, Greece
[2]School of Electrical and Computer Engineering, NTUA, Greece

Resnik (1996)

Selectional constraints verbs that strongly select their objects, such as *eat* and *drink*, tend to drop them as easily recoverable.

Alternating Verbs	Non – Alternating Verbs
call	bring
eat	catch
drink	do
explain	find

Levin (2000)

The Argument per Sub-event Condition: There must be at least one argument XP in the syntax per sub-event in the event structure.
Cause: Two sub-events → Obligatory objects
Activity-Accomplishment: One sub-event → Non obligatory objects

Ruppenhoffer (2005) – Goldberg (2005)
Object ommision: lexically specified; induced by certain grammatical constructions

Hopper – Thompson (1980)

Transitivity: not an on-off property.

Materials
• Hellenic National Corpus (~50M words)
• 28 verbs in the semantic domains of ingestion, natural senses and contact (Th. Vostantzoglou. 1962. *Onomastikon*); average 487 sentenes/verb
• 54 Senses; average 246 sentences /verb

Method
• **Annotate and group sentences according to sense**, drawing on own intuitions without reference to a lexicon
• **Grouping criteria:** Synonymy; Translation to English; Zeugma; Subcategorisation properties; Participation to certain verb alternations
• **Empirical grounding:** For 10 polysemous verbs, **inter-annotation** and **inter-grouping** experiments with 20 native speakers. We studied how and achieved to:
 • **Maximise** degree of **community acceptance** for verb sense grouping
 • **Maximise inter-annotator agreement**
• Crucial: **Set apart** Multi-Word Expressions and Light Verb Constructions **(Rumshisky & Batiukova, 2008)**
• Distinguish between **primary** and **secondary senses** on native speaker knowledge and frequency of use
FURTHERMORE:
• Label senses according to aspectual properties:
 • ACH - Achievements/Accomplishments: Events with inherent telic point
 • ACT - Activities: Event with no inherent telic point.
• Identify causative and non-causative verbal senses

Results

Senses Clustering:

$$X - axis : \frac{\#trans.occurences}{\#occurences}$$

$$Y - axis : \frac{\#clitics}{\#trans.occurences}$$

1. Causatives tend to cluster together (**Fig. 1**)
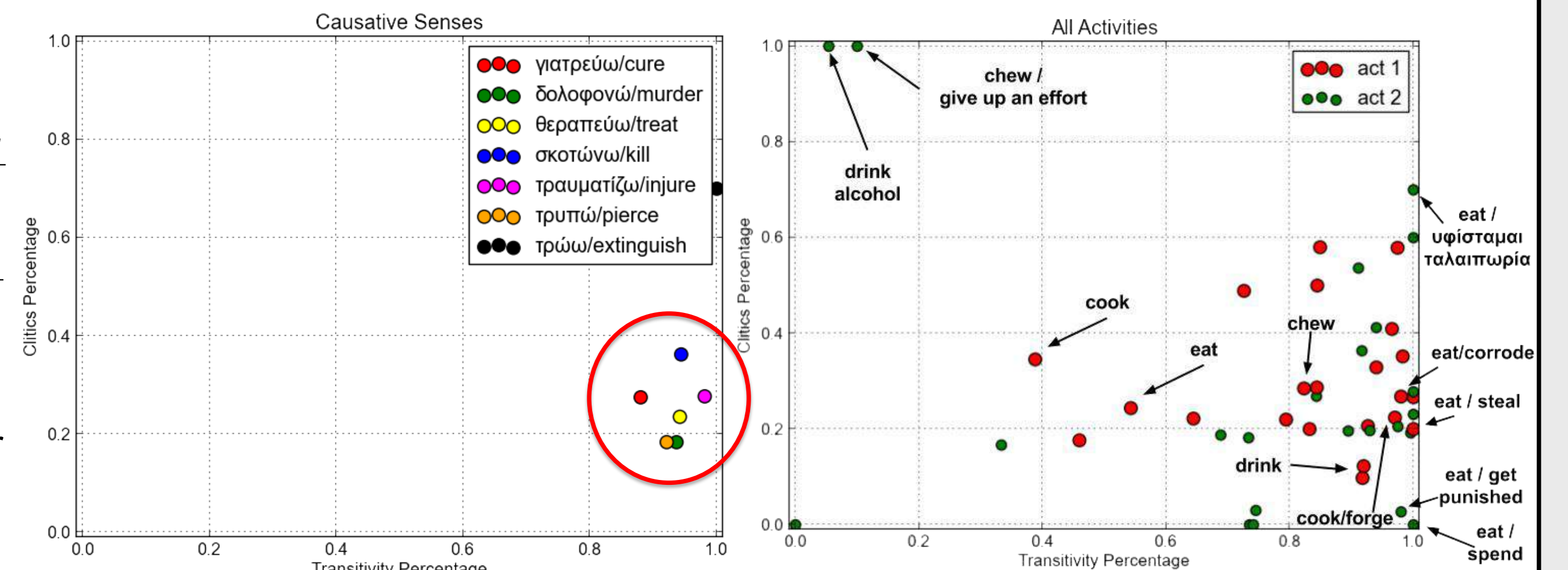2. Activities require an obligatory object only in systems (**Fig. 2**)

Fig. 1 Causatives

Fig. 2 Activities and Polysemy

3. Different senses of the same verb tend to be far apart (Fig. 2)

Interannotation:
Round 1: all possible senses, including LVCs and MWEs.
Round 2: all possible senses, LVCs and MWEs are excluded, general rather than specific senses eg. ξοδεύω > σπαταλώ.

Grouping: no senses given.

Metrics:
1. Rand Index: Partition similarity (Table 1)
2. Inter-annotation aggreement: pairwise average (Table 2)

Comparison with established lexica

Ρήματα	Inter-annotation Round 1	Inter-annotation Round 2	Grouping
Αγγίζω	0.7873	0.8021	0.8456
Ακούω		0.6281	0.6140
Αντιλαμβάνομαι	0.5689		
Αρπάζω	0.6794	0.834	0.7802
Δοκιμάζω	0.7942		
Καταβροχθίζω	0.76		
Κοιτάζω		0.7049	0.7339
Μυρίζω		0.6921	0.7746
Τρώω	0.9105	0.955	0.9109

Table 1. Rand Index

Ρήματα	Inter-annotation Round 1		Inter-annotation Round 2	
	Mean	StDev	Mean	StDev
Αγγίζω	0.34	0.0712	0.745	0,0762
Ακούω			0.4667	0,2082
Αντιλαμβάνομαι	0.68	0.1619		
Αρπάζω	0.3962	0.1357	0.7448	0.0878
Δοκιμάζω	0.6368	0.1252		
Καταβροχθίζω	0.66	0.0966		
Κοιτάζω			0.5652	0
Μυρίζω			0.4921	0.055
Τρώω	0.5375	0.0766	0.8595	0.0398

Table 2. Interannotator agreement

Λεξικό της κοινής νεοελληνικής	Our senses
Ακουμπώ κτ ή κπ. με το χέρι	Ακουμπώ
Δοκιμάζω	
Οικειοποιούμαι	
Φτάνω κάπου, προσεγγίζω	Φτάνω
Αφορώ	Έχω κάποια επίδραση σε
Συγκινώ	Συγκινώ
Πειράζω, ενοχλώ	🚫
🚫	Is nearly

Table 3. Senses of αγγίζω

Λεξικό της κοινής νεοελληνικής	Our senses
Μασώ και καταπίνω στερεά ή ημίρρευστη τροφή	Εσθίω
Προκαλώ σε κτ υλική φθορά	Φθείρω
Ξοδεύω ή σπαταλώ κτ	Ξοδεύω,σπαταλώ
Προξενώ σε κτ σωματική ή ψυχική φθορά	Προξενώ σωματική επίθεση-ταλαιπωρία
Προξενώ σε κάποιον κακό	
Δεχομαι ποινή	Τιμωρούμαι
Κλέβω ή ιδιοποιούμαι	Κλέβω
Ανέχομαι, υπομένω	υφίσταμαι
Για κάτι που πέφτει επάνω μου	
Έχω φαγούρα	Έχω φαγούρα
Δείχνω προτίμηση για φαγητό	
Αφαιρώ τμήμα από υλικό	🚫
Ξεγελιέμαι, εξαπατώμαι	
Ζητώ, επιδιώκω επίμονα	

Table 4. Senses of τρώω

Conclusions

A. Corpora annotated with generally accepted verb senses:

Inter-annotation scores are significantly improved when the "noisy" Multi-Word Expressions (MWEs) and Light Verb Constructions (LVCs) are removed from the data.
Inter-grouping experiments produce comparable inter-annotator agreement results with the stricter inter-annotation experiments of the 2nd round. Grouping of sentences on the basis of the sense of their verb head is certainly a less biased procedure than pre-specified label assignment (Kilgarriff, 1997). It is not clear how work load is divided between the annotators and the experimenters with each of these methods.

B. The phenomenon of transitivity:

Transitivity is not an on-off phenomenon. 'Degrees of transitivity' seem to help differentiate between different senses of a "transitive" verb, ie. a verb that accepts an object. Perhaps, "degree of transitivity" is part of the sub-categorisation properties of a verb lemma.

'Degree of transitivity' is depicted here as the combined result of object presence and usage of clitics. To the best of our knowledge, this is the first time that clitics distribution is used as a trait of transitivity.

Future work
Enrich the data and the observations of this research in order to develop a system for VSD for Modern Greek.

Literature

Βοσταντζόγλου, Θεολόγος. 1962. *Αντιλεξικόν ή Ονομαστικόν της Νεοελληνικής Γλώσσης*, Αθήνα.

Cruse, D.A. 1991. *Lexical Semantics*. Cambridge Textbooks in Linguistics, Cambridge University Press.

Goldberg, Adele E. 2005. 'Constructions, Lexical Semantics and the Correspondence Principle: Accounting for Generalizations and Subregularitiesin the Realization of Arguments'. Nomi Erteschik-Shir and Tova Rapoport (eds.) *The Syntax of Aspect* Oxford University Press.

Hopper, Paul J. and Thompson, Sandra A. 1980. 'Transitivity in Grammar and Discourse.' *Language*, Vol. 56, No. 2 (Jun., 1980): 251-299.

Kilgarriff, Adam. 1997. 'I don't believe in word senses'. *Computers and the Humanities* 31: 91-113.

Eric Laporte, Elisabete Ranchhod, Anastasia Yannacopoulou. 'Syntactic variation of support verb constructions.' *Lingvisticae Investigationes*, 2008, 31 (2): 173-185.

Levin, Beth. 2000. 'Aspect, Lexical Semantic Representation, and Argument Expression'. In Lisa J. Conathan et al. (eds.): *Proceedings of the 26th Annual Meeting of the Berkeley Linguistics Society*, Berkeley : 413-429.

Resnik, Philip. 1996. 'Selectional constraints: An information-theoretic model and its computational realization'. *Cognition* 61: 127-159.

Rumshisky, Anna and Batiukova, Olga. 2008. "Polysemy in verbs: systematic relations between senses and their effect on annotation". *Coling 2008: Proceedings of the workshop on Human Judgements in Computational Linguistics*, August 2008, Manchester, UK: 33-41.

Ruppenhofer, Josef. 2005. *Regularities in null instantiation*.

Yager, Ronald R. 2008. 'Some measures relating partitions useful for Computational Intelligence'. *International Journal of Computational Intelligence Systems*, *Vol.1, No. 1* (January, 2008): 1–18, Atlantis Press.