

# Phonology Competes with Syntax: Experimental Evidence for the Interaction of Word Order and Accent Placement in the Realization of Information Structure

Frank Keller and Theodora Alexopoulou

Division of Informatics and

Department of Theoretical and Applied Linguistics

University of Edinburgh

2 Buccleuch Place, Edinburgh EH8 9LW, UK

phone: +44-131-650-4437, fax: +44-131-650-6626

email: keller@cogsci.ed.ac.uk, theodora@ling.ed.ac.uk

Final Version, November 8, 2000

## Abstract

In this paper, we investigate the interaction of phonological and syntactic constraints on the realization of Information Structure in Greek, a free word order language. We use magnitude estimation as our experimental paradigm, which allows us to quantify the influence of a given linguistic constraint on the acceptability of a sentence. We present results from two experiments. In the first experiment, we focus on the interaction of word order and context. In the second experiment, we investigate the additional effect of accent placement and clitic doubling. The results show that word order, in contrast to standard assumptions in the theoretical literature, plays only a secondary role in marking the Information Structure of a sentence. Order preferences are relatively weak and can be overridden by constraints on accent placement and clitic doubling. Our experiments also demonstrate that a null context shows the same preference pattern as an all focus context, indicating that 'default' word order and accent placement (in the absence of context) can be explained in terms of Information Structure.

In the theoretical part of this paper, we formalize the interaction of syntactic and phonological constraints on Information Structure. We argue that this interaction is best captured using a notion of grammatical competition, such as the one developed by Optimality Theory (Prince & Smolensky, 1993, 1997). In particular, we exploit the optimality theoretic concept of constraint ranking to account for the fact that some constraint violations are more serious than others. We extend standard Optimality Theory to obtain a grammar model that predicts not only the optimal (i.e., grammatical) realization of a given input, but also makes predictions about the relative grammaticality of suboptimal structures. This allows us to derive a constraint hierarchy that accounts for the interaction of phonological and syntactic constraints on Information Structure and models the acceptability patterns found in the experimental data.

## 1. Introduction

Information Structure refers to an independent level of grammar encoding the *packaging* of the linguistic message in order to meet discourse requirements. A diverse set of theoretical primitives has been proposed to conceptualize Information Structure, including notions such as *given/new*, *ground/focus*, or *theme/rheme* (see Vallduví & Engdahl, 1996, for a review). In this paper, we follow the approach of Vallduví (1992), which assumes that a sentence is partitioned into focus and ground, with ground further subdivided into link and tail (see below for details).

There is considerable crosslinguistic variation as to the grammatical devices that languages employ to mark Information Structure. English, for instance, mainly exploits phonological markers (accent placement), while languages such as Catalan, Finnish, Greek, or Hungarian have been argued to rely mainly on word order (Kiss, 1995b; Horvath, 1995; Tsimpli, 1995; Vallduví, 1995; Vilkuna, 1995).

In the present paper we focus on Greek, a language employing both phonological and syntactic means (word order and clitic doubling) for the realization of Information Structure, and provide experimental data on the interaction of these factors. Our aim is to integrate previous experimental work, which has either focussed on the phonological manifestation of Information Structure (e.g., Birch & Clifton, 1995), or on syntactic factors (e.g., clefting, Vion & Colas, 1995). We use magnitude estimation (Bard, Robertson, & Sorace, 1996; Cowart, 1997) to obtain acceptability judgments for contextualized sentences. Magnitude estimation has been shown to yield maximally fine-grained measurements of linguistic intuitions, while avoiding the pitfalls of conventional, ordinal judgment scales (Cowart, 1997; Schütze, 1996; Sorace, 1992). The results of these magnitude estimation experiments allow us to quantify the contribution of accent placement, word order, and clitic doubling to the information structural well-formedness of a sentence.

In the theoretical part of this paper, we attempt to formalize the interaction of syntactic and phonological constraints on Information Structure. We argue that this interaction is best captured using a notion of grammatical competition, such as the one developed by Optimality Theory (Prince & Smolensky, 1993, 1997). In particular, we exploit the optimality theoretic concept of constraint ranking to account for the fact that some constraint violations are more serious than others. We extend standard Optimality Theory to obtain a grammar model that predicts not only the optimal (i.e., grammatical) realization of a given input, but also makes predictions about the relative grammaticality of suboptimal structures. This allows us to derive a constraint hierarchy that accounts for the interaction of phonological and syntactic constraints on Information Structure and models the acceptability patterns found in the experimental data.

This paper is organized as follows. In the remainder of this section, we give an overview of the basic facts regarding Information Structure in English and Greek. Section 2 presents the theoretical framework that we use to account for the interaction of syntactic and phonological constraints

---

Thanks to Ellen Bard, Caroline Féry, Bob Ladd, Ineke Mennen, Maria Lapata, Scott McDonald, Caroline Heycock, Mark Steedman, Enric Vallduví and four anonymous reviewers for comments. We are particularly grateful to Vasilios Karaiskos for reading the stimuli and to Christina Anagnostopoulou, Maria Lapata, and Ineke Mennen for taking part in the pretests. The first author acknowledges the support of an ESRC Postgraduate Research Studentship. The second author is grateful to Richard Coates and the Linguistics Group at the School of Cognitive and Computing Sciences, University of Sussex, for organizing a visiting fellowship. Part of the work reported in this paper was completed during this fellowship at Sussex.

on Information Structure. We first introduce standard Optimality Theory, and then propose an extended version of Optimality Theory that employs a relative notion of grammaticality and allows us to model degrees of acceptability. We conclude Section 2 by describing the set of optimality theoretic constraints we assume, based on the linguistic background presented in Section 1. Sections 3 and 4 describe two experiments designed to clarify the competition of syntactic and phonological constraints on Information Structure. The first experiment deals with written stimuli and focuses on the role of word order as a marker of Information Structure. The second experiment uses speech stimuli and investigates the interaction of word order, accent placement, and clitic doubling. At the end of Sections 3 and 4, we derive constraint rankings from the experimental data and discuss how well these rankings model the acceptability patterns found experimentally. Finally, Section 5 draws some general conclusions and provides an outlook on further research.

### 1.1. Information Structure in English

This section gives an overview of the concepts used in the Information Structure framework proposed by Vallduví (1992). Building on previous work of, among others, Chafe (1976, 1983), and Prince (1986), Vallduví (1992) views a sentence as conveying information that updates the hearer's knowledge-base or *information state*. Each sentence constitutes an *instruction* indicating to the hearer *what* information to add, *where* to add it, and *how*. These instructions are encoded in the *Information Structure* of a sentence, which consists of the following primitives:

- (1) Sentence := {Focus, Ground}  
 Ground := {Link, Tail}

*Focus* conveys the new information of the sentence, whereas *ground* anchors the new information to the hearer's current information state. Ground is further subdivided into *link* and *tail*: *link* points to the locus of update in the hearer's information state, i.e., to where the new information should be added. *Tail* indicates *how* information should be added.<sup>1</sup> The three primitives *focus*, *link*, and *tail* combine to yield four instruction types: *all focus*, *link-focus*, *focus-tail*, and *link-focus-tail*. Below we explicate briefly the function of each of these instruction types. Possible realizations of the four instructions in English are exemplified in (2); SMALL CAPITALS indicate main sentential stress, **boldface** marks secondary stress. Focus is indicated using square brackets and subscript F.

- (2) a. **All Focus**  
 The president has a weakness.  
 [F He hates CHOCOLATE].
- b. **Link-Focus**  
 Tell me about the people in the White House. Anything I should know?  
 The **president** [F hates CHOCOLATE].
- c. **Focus-Tail**  
 You shouldn't have brought chocolates for the president.  
 [F He HATES] chocolate.

<sup>1</sup>Vallduví's tripartite organization of Information Structure combines previous distinctions such as *theme-rheme*, *topic-comment*, and *ground-focus* (Halliday, 1967; Reinhart, 1982). In particular, his *link* corresponds to traditional notions of topic or theme.

d. **Link-Focus-Tail**

And what about the president? How does HE feel about chocolate?

The president [<sub>F</sub> HATES] chocolate.

(Vallduví & Engdahl, 1996)

All instruction types contain a focus part, since every sentence has an update potential. The presence of ground segments depends on the knowledge shared between the interlocutors in the previous discourse, i.e., on the context in which the sentence is uttered. Let us consider the above examples in more detail. Sentence (2a) involves an all focus instruction which updates the information about *the president*. Note that *the president* has already been activated as a locus of update by the context in which (2a) appears. Hence (2a) does not contain a link.<sup>2</sup> Similarly, the locus of update is inherited from the previous context in (2c), which also conveys a ‘link-less’ instruction. On the other hand, example (2b) specifies *the president* as its link, i.e., the locus of update. (2b) instructs the hearer to add the new condition *hates chocolate* to this locus. Finally, example (2d) also specifies *the president* as the locus of update, but conveys a different update instruction (tail). It instructs the hearer to search for a condition of the form *likes chocolate* and replace the predicate *likes* with *hates*. Note that the same instruction is also encoded by the tail in (2c).

Let us turn to the linguistic means by which English marks the different components of Information Structure. All of the examples in (2) involve the same word order. However, they differ in their intonational pattern: English relies on prosodic means for encoding Information Structure. Focused segments are associated with the main sentential stress (A accent). This is true of *narrow focus* as in (2d), but also of *broad focus* as in (2b). (Narrow and broad focus are descriptive terms: narrow focus denotes NP or verb focus, while broad focus refers to VP or S focus). Any accented constituent can be interpreted as narrow focus, while only accent on the rightmost complement can give rise to a broad focus interpretation (Vallduví, 1992; Ladd, 1996).<sup>3</sup> English provides intonational marking not only for foci, but also for links. Links like *the president* in (2b) receive secondary stress. (This accent is referred to as B accent in the theoretical literature, see Ladd, 1996; Pierrehumbert & Hirschberg, 1990; Steedman, 1991. For further discussion on the realization of Information Structure in English see also Bolinger, 1978, 1989; Selkirk, 1984; Rochemont, 1986.)

Note that our choice of the framework of Information Structure does not bear directly on the claims we make in subsequent sections. It mainly serves as a theoretical background against which we discuss the phenomena at hand. However, Vallduví’s assumption that Information Structure forms an independent grammatical level, interacting with both syntax and phonology, is compatible with the model of constraint interaction we will advocate in the remainder of this paper. Note further that the experiments we will report only investigate the ground-focus distinction and do not explore the distinction between link and tail. This restriction was necessary to keep the experimental design to a manageable size. For completeness, the present section introduced all three information structural primitives.

### 1.2. Information Structure in Greek

As in English, accent placement plays a central role in the realization of Information Structure in Greek. However, in addition to phonological resources, Greek also employs syntactic resources

<sup>2</sup>Pronouns do not contribute to the Information Structure of a sentence. They are just syntactic placeholders (Vallduví, 1995).

<sup>3</sup>Technically, it is the most oblique rather than the rightmost NP in English (Vallduví & Engdahl, 1996). However, the most oblique NPs in English are typically the rightmost ones.

such as word order and clitic doubling to realize Information Structure. In this section, we briefly present how phonological and syntactic devices combine to yield various Information Structure instructions in Greek.

Throughout this paper, we will employ a notation that uses capitalization to indicate accent, e.g., svO indicates the order subject-verb-object with accent on the object. The same order with clitic doubling is denoted as sclvO. We use an all capital notation where we disregard accent, such as in SVO.

Let us first consider cases of narrow focus as in (3) where the subject NP *o Yanis* is focused. Focused NPs are accented, as in English, but, unlike English, their order is not fixed; they may appear either preverbally (see (3b)) or postverbally (see (3a)). (According to the literature, preverbal focus is more likely to be associated with a contrastive reading and therefore might be considered slightly more marked than postverbal focus, see Alexopoulou, 1998; Tsimpli, 1995.) Not only focused NPs that are free to appear either preverbally or postverbally; so are ground NPs (*ti Maria* in (3)). The standard literature assumes that preverbal ground NPs realize Vallduvian links (or topics in the traditional sense), while postverbal ones are interpreted as tails (Alexopoulou, 1998; Anagnostopoulou, 1994; Philippaki-Warbuton, 1985; Schneider-Zioga, 1994; Tsimpli, 1995; Valioui, 1994). Hence example (3a) realizes a *link-focus* instruction in Vallduví's terms, while (3b) realizes a *focus-tail* instruction.

(3) **Subject Focus**

Pios apelise ti Maria?

'Who fired Maria?'

- a. Ti Maria tin apelise [<sub>F</sub> o YANIS].  
 the Maria-ACC her-CL fired-3SG [<sub>F</sub> the Yanis-NOM]  
 'Yanis fired Maria.'
- b. [<sub>F</sub> O YANIS] (tin) apelise ti Maria.  
 [<sub>F</sub> the Yanis-NOM] her-CL fired-3SG the Maria-ACC

Both ground and focused NPs can alternate between preverbal and postverbal positions. Ground NPs, however, tend to appear in peripheral positions while focused ones are preferred adjacent to the verb (Alexopoulou, 1998; Schneider-Zioga, 1994; Tsimpli, 1995). Thus, the clvSo order in (4b), where the focused subject NP is adjacent to the verb, is a felicitous answer to (3); in contrast, the clvoS sentence in (4a), where the focused NP is dislocated to the right periphery of the clause, is infelicitous.

(4) **Subject Focus**

Pios apelise ti Maria?

'Who fired Maria?'

- a. ??Tin apelise ti Maria [<sub>F</sub> o YANIS].  
 her-CL fired-3SG the Maria-ACC [<sub>F</sub> the Yanis-NOM]  
 'Yanis fired Maria.'
- b. Tin apelise [<sub>F</sub> o YANIS] ti Maria.  
 her-CL fired-3SG [<sub>F</sub> the Yanis-NOM] the Maria-ACC

Similarly, preverbal ground NPs typically precede preverbal focus (see (5a)). (In fact, various authors consider examples like (5b), where the link follows the focused NP, ungrammatical, see Tsimpli, 1995; Tsipakou, 1998.)

- (5) a. Ti Maria [F O YANIS] tin apelise horis kamia proidopiisi.  
 the Maria-ACC [F the Yanis-NOM] her-CL fired-3SG without any warning  
 ‘Yanis fired Maria without any warning.’
- b. ??[F O YANIS] ti Maria tin apelise horis kamia proidopiisi.  
 [F the Yanis-NOM] the Maria-ACC her-CL fired-3SG without any warning

Note finally, that the ground object NP in (3) is marked by an additional clitic pronoun, attached to the verb. We will use the term *clitic doubling* to refer to a configuration where the object NP is co-indexed with such a clitic, irrespective of the position of the object NP.<sup>4</sup> When the NP is postverbal, doubling is optional (as indicated by the brackets around the clitic *tin* in (3b)); but doubling tends to be obligatory when the NP is dislocated to the left (we return to this issue later in this section).

The ground-focus partition is thus realized in Greek through the exploitation of diverse structural resources: accent placement (on the focused constituent), word order (focused NPs are adjacent to the verb, ground ones are dislocated to peripheral positions), and clitic doubling (object ground NPs are preferred doubled). The interaction between these structural devices follows a consistent pattern, independent of the grammatical function of the focused or ground NP (modulo the fact that clitic doubling is only available for objects in Greek). Consider the examples in (6) which demonstrate a narrow focus reading for the object NP:

(6) **Object Focus**

Pion apelise i Maria?

‘Who did Maria fire?’

- a. I Maria apelise [F to YANI].  
 the Maria-NOM fired-3SG [F the Yanis-ACC]  
 ‘Maria fired Yanis.’
- b. [F To YANI] apelise i Maria.  
 [F the Yanis-ACC] fired-3SG the Maria-NOM
- c. \*[F To YANI] ton apelise i Maria.  
 [F the Yanis-ACC] him-CL fired-3SG the Maria-NOM

Again, the accent falls on the focused NP *to Yani*, which can appear either preverbally (see (6b)) or postverbally (see (6a)), while the ground subject NP *i Maria* is unaccented. Again, the focused object NP is preferred adjacent to the verb, as indicated by the felicity of the vOs order in (7a) compared to the vsO order in (7b), as answers to the question in (7).

(7) **Object Focus**

Pion apelise i Maria?

‘Who did Maria fire?’

- a. Apelise [F to YANI] i Maria.  
 fired-3SG [F the Yanis-ACC] the Maria-NOM  
 ‘Maria fired Yanis.’
- b. ??Apelise i Maria [F to YANI].  
 fired-3SG the Maria-NOM [F the Yanis-ACC]

Focused objects cannot be doubled, as the unacceptability of (6c) indicates. In fact, sentences with accent on a clitic doubled object are unacceptable, irrespective of the context they appear in (Agouraki, 1993; Alexopoulou, 1998), as illustrated by (8):

<sup>4</sup>Our term subsumes clitic doubling and clitic left dislocation which are used for postverbal and left dislocated clitic doubled NPs, respectively.

- (8) a. \*To YANI ton ida.  
           the Yani-ACC him-CL saw-1SG  
           ‘I saw Yanis.’  
 b. \*Ton ida to YANI.  
           him-CL saw-1SG the Yani-ACC

The general unacceptability of accented doubled objects is due to the conflicting information structural requirements imposed on these objects. Accent marks the object as focus, while doubling marks it as ground.

Verb focus is marked with accent on the verb as in (9). Again, the ground NPs can appear either preverbally (see (9a)) or postverbally (see (9b)), while the object NP is preferred doubled:

- (9) **Verb Focus**  
 Ti ekane o Yanis me to aftokinito?  
 ‘What did Yanis do with the car?’  
 a. O Yanis [F to PULISE] to aftokinito.  
    the Yanis-NOM [F it-CL sold-3SG] the car-ACC  
    ‘Yanis sold the car.’  
 b. To aftokinito [F to PULISE] o Yanis.  
    the car-ACC [F it-CL sold-3SG] the Yanis-NOM

Let us now turn to all focus instructions. VSO is standardly considered the most natural response to an all focus question (see (10)). In fact, the naturalness of VSO in this context has been part of the argument for the standard analysis of VSO as the basic order of Greek (Agouraki, 1993; Alexopoulou, 1998; Philippaki-Warburton, 1985; Tsimpli, 1995). In such contexts, accent falls on the rightmost constituent. In this respect, Greek seems to pattern with English; in both languages an all focus interpretation arises from accent on the rightmost NP.

- (10) **All Focus**  
 Kana neo?  
 ‘Any news?’  
 [F pulise o Yanis to AFTOKINITO].  
 [F sold-3SG the Yanis-NOM the car-ACC]  
 ‘Yanis sold the car.’

It is worth mentioning here that, as Vallduví and Engdahl (1996) note, questions introducing an all focus context (*What happened/Any news?*) can also give rise to VP focus, with the subject or object dislocated to the left periphery of the clause. Indeed, svO (see (11a)) and oclvS (see (11b)), instantiating a link/topic-focus Information Structure, are also felicitous answers to an all focus question.

- (11) **All Focus**  
 Kana neo?  
 ‘Any news?’  
 a. O Yanis pulise to AFTOKINITO.  
    the Yanis-NOM sold-3SG the car-ACC  
    ‘Yanis sold the car.’  
 b. Tis fises tha tis stilume AVRIO.  
    the posters-ACC FUT them-CL send-1PL tomorrow  
    ‘We will send the posters tomorrow.’

Broad and narrow focus contexts differ significantly in the range of utterances they can accommodate. A broad focus context allows the accommodation of a wider range of ground-focus partitions, while a narrow focus context only accepts sentences with a ground-focus partition strictly corresponding to the expectations it imposes. For instance, example (12) can be a felicitous answer to a question like (10), even though it does not directly correspond to an all focus instruction; rather than the rightmost constituent, the accent falls on the verb, while the object NP is doubled. This sentence is acceptable in a context where the two interlocutors share the knowledge that Yanis was expected to sell his car. However, even if such knowledge is shared by the two speakers, such a sentence would not be acceptable as an answer to an object focus question like *What did Yanis sell?*; only a ground (object) focus instruction constitutes a felicitous answer for this question.

- (12) To PULISE to aftokinito o Yanis.  
 it-CL sold-3SG the car-ACC the Yanis-NOM  
 ‘Yanis sold the car.’

It is worth mentioning here that the wider range of answers satisfying an all focus question yields higher freedom in the linguistic realization of these answers. Thus, most orders (SVO, OVS, and the verb initial orders) are acceptable, while the accent may shift from the rightmost clause boundary to the left.

To summarize, Information Structure in Greek is realized through a combination of phonological and syntactic means, captured by the following descriptive generalizations:

(13) **Descriptive Generalizations on Information Structure in Greek**

- a. **Phonology:** (i) Accented constituents are (part of) focus; ground elements bear no accent; (ii) accent on the rightmost NP gives rise to a broad focus interpretation.
- b. **Word Order:** ground constituents are peripheral.
- c. **Clitic Doubling:** doubled objects are ground.

Note that, while it is true that ground NPs are peripheral, it is not always the case that focused NPs are adjacent to the verb. Adjacency is observed in cases of narrow focus but not always in all focus instructions. In these instructions accent falls on the rightmost NP which, very often, is not adjacent to the verb (e.g., vsO or voS).

In the following, we will briefly comment on two more restrictions on word order in Greek. First, as mentioned earlier, preverbal ground objects (see (3a)) should be doubled, while doubling is optional with postverbal ground NPs. The obligatoriness of the clitic in examples like (14) has been a matter of controversy in the literature. Examples like these are often judged less acceptable when they lack doubling, with some authors judging them unacceptable (Tsiplakou, 1998). One goal of the experimental study in this paper is to settle the data disputes that surround these examples.

- (14) a. To aftokinito ??(to) pulise o YANIS.  
 the car-ACC it-CL sold-3SG the Yanis-NOM  
 ‘Yanis sold the car.’
- b. Tin Maria ??(tin) apelise o YANIS.  
 the Maria-ACC her-CL fired-3SG the Yanis-NOM  
 ‘Yanis fired Maria.’

In Section 2.3 we will postulate a constraint requiring preverbal objects to be doubled and we will provide evidence supporting this constraint in Section 4.



The second issue concerns verb final word orders. So far we have only considered orders in which the verb appears in either initial or medial position. Verb final orders (SOV and OSV), while grammatical, are generally perceived as less acceptable:

- (15) a. ??Tis Marias o Yanis / o Yanis tis Marias tis milise.  
 the Maria-GEN the Yanis-NOM / the Yanis-NOM the Maria-GEN her-CL talked-3SG  
 ‘Yanis talked to Maria.’  
 b. ??Ta pedia o Yanis / o Yanis ta pedia ta ide.  
 the kids-ACC the Yanis-NOM / the Yanis-NOM the kids-ACC them-CL saw-3SG  
 ‘Yanis saw the kids.’

However, SOV and OSV improve to full acceptability if more material is added after the verb:

- (16) a. Tis Marias o Yanis / o Yanis tis Marias tis ipe oti de theli  
 the Maria-GEN the Yanis-NOM / the Yanis-NOM the Maria-GEN her-CL said-3SG that not want  
 na pai stin Ameriki.  
 SUBJ go to-the America  
 ‘Yanis told Maria that he doesn’t want to go to America.’  
 b. Ta pedia o Yanis / o Yanis ta pedia ta vlepi mono otan den  
 the kids-ACC the Yanis-NOM / the Yanis-NOM the kids-ACC them-CL see-3SG only when not  
 ehi dulia to Savatokiriako.  
 have work the weekend  
 ‘Yanis sees the kids only when he doesn’t have work during the weekend.’

To account for the reduced acceptability of (15) compared with (16), we will assume a constraint that penalizes verbs that occur at clause final positions (see Section 2.3).

To summarize: all of the factors discussed in this section are expected to have a significant effect on the acceptability of a given word order. This includes the information structural factors listed in (13), as well as the restrictions on preverbal objects and on verb final sentences stated above. However, it quickly becomes evident that not all of these factors play an equally important role. One of the main goals of the experiments reported in the remainder of this paper is to identify the nature of the interaction between these factors and to quantify the effect of each of them. Before we present the experimental results, some preliminary observations are in order.

Accent placement appears as the most important factor in the realization of the ground-focus partition as it is both obligatorily and unambiguously associated with (at least part of) focus. Word order, on the other hand, appears as a comparatively weak factor. In the absence of accent and clitic doubling, a given order may give little or no indication of the ground-focus partition; for example, SVO and OVS can realize a link-focus or focus-ground partition, depending on accent placement and doubling (svO/oclvS and Sclvo/Ovs). Similarly, VSO can realize an all focus sentence or allow a narrow focus interpretation for the subject NP (vsO and clvSo respectively). Unlike word order, clitic doubling is unambiguously associated with a ground interpretation of objects. However, unlike accent, doubling is not necessary for the realization of ground NPs, and its effect is restricted to objects.

In Section 2.3, we will introduce a set of grammatical constraints based on the generalizations presented above. We expect that the experimental results will show that all these constraints play a role in the realization of Information Structure, while the size of the acceptability difference caused by each of these factors will reflect its relative importance. More precisely, we expect violations of accent placement to induce the strongest effect. Given its unambiguous association with focus,

accent placement provides hearers with a strong cue for the Information Structure of a sentence. The restriction that doubled NPs cannot function as foci is also expected to produce strong effects. Just as accent placement, clitic doubling is an unambiguous marker of Information Structure. Violations of word order preferences, on the other hand, are expected to trigger weak effects; given its ambiguity, word order is an additional, but rather unreliable cue for detecting the ground-focus partition of a sentence. Note as well that, due to the ambiguity of word order, some word orders will satisfy the information structural requirements of several contexts.

## 2. Modeling Gradience and Grammatical Competition

One of the main goals of this study is to investigate the *relative* contribution of the factors that play a role in the realization of Information Structure in Greek, namely word order, accent placement, and clitic doubling. As discussed at the end of the previous section, these three factors do not induce equally strong effects on acceptability, but give rise to *gradient* acceptability judgments. Our hypothesis is that such gradience is the result of differences in the *strength* of the constraints that govern order, accent, and doubling. It follows that a model that captures gradient data adequately has to (a) provide a way of accounting for relative constraint strength, and (b) based on this notion of constraint strength, offer a gradient notion of grammaticality. The predictions made by such a model can then be tested against gradient acceptability data obtained experimentally.

In order to satisfy these theoretical desiderata, we adopt a model that is based on the notion of *constraint ranking* and makes predictions about the relative rather than absolute grammaticality of linguistic structures. The model builds on core concepts from Optimality Theory (OT; Prince & Smolensky, 1993, 1997), a linguistic framework which enjoys considerable popularity in phonology (and to a lesser extent in syntax). Optimality Theory is attractive for our purposes as it is equipped with a notion of constraint competition that allows us to describe the interaction of linguistic constraints (in our case syntactic and phonological constraints on Information Structure). Furthermore, OT provides a notion of constraint ranking that allows us to account for the fact that constraints differ in strength, i.e., that some constraints are more important than others for the overall well-formedness of a given linguistic structure.

While OT lends itself to a flexible account of the interaction of syntactic and phonological constraints, and provides a notion of constraint ranking that we can build on, it does not offer a relative notion of grammaticality. On a par with traditional generative linguistics, standard OT is based on the assumption that grammaticality is a binary concept. To overcome this limitation, we will propose an extended version of OT, based on the *suboptimality hypothesis*. This hypothesis leads to a relative notion of grammaticality that covers both optimal and suboptimal linguistic structures.

In the following, we will first (in Section 2.1) give a brief overview of the concepts of standard OT, as proposed by Prince and Smolensky (1993, 1997). We will then introduce our modified version of OT in Section 2.2. The particular constraints we assume for Information Structure in Greek are introduced in Section 2.3.

### 2.1. Standard Optimality Theory

Standard Optimality Theory deviates from more traditional linguistic frameworks in that it assumes grammatical constraints to be (a) universal, (b) violable, and (c) ranked. Assumption (a) means that constraints are maximally general, i.e., they contain no exceptions or disjunctions, and there is no parameterization across languages. Highly general constraints will inevitably conflict;

therefore assumption (b) allows constraints to be violated, even in a grammatical structure, while assumption (c) states that some constraint violations are more serious than others. While, according to (a), the formulation of constraints remains constant across languages, the ranking of the constraints can differ between languages, thus allowing us to account for crosslinguistic variation.

In an OT setting, a structure is grammatical if it is the *optimal* structure in a set of candidate structures. Optimality is defined via constraint ranking: the optimal structure violates the least highly ranked constraints compared to its competitors. The number of violations plays a secondary role; if two structures violate a constraint with the same rank, then the number of violations incurred decides the competition. OT therefore deviates from traditional grammatical frameworks in that the grammaticality of a sentence is not determined in isolation, but in comparison with other possible structures. Note that there is no inherent restriction on the number of optimal candidates for a given candidate set; more than one candidate may be optimal if several candidates share the same constraint profile, i.e., if they incur the same constraint violations.

We will illustrate how OT works with a simple example taken from an account of *wh*-extraction by Legendre, Wilson, Smolensky, Homer, and Raymond (1995). Our example deals with extraction from direct objects in English. Legendre et al. (1995) assume that the following constraints govern extraction: SUBCAT, which states that the subcategorization requirements of the verb have to be met; \*t, which disallows traces (i.e., movement); and BAR $n$ , which rules out movement that crosses more than  $n$  barriers (for a definition of barrier, see Legendre et al., 1995). For English, the assumption is that these constraints are ranked as follows:

(17) SUBCAT  $\gg$  BAR4  $\gg$  BAR3  $\gg$  BAR2  $\gg$  \*t

This means that a violation of SUBCAT is more serious than a violation of BAR4, which in turn is more serious than a violation of BAR3, etc.

A crucial assumption in OT is that all candidate structures (syntactic representations) that take part in a grammatical competition are generated from a common input, assumed to be a predicate argument structure by Legendre et al. (1995). The input structure specifies the verb and the arguments of the verbs, plus operators and scope relations that might be present. As an example, consider the first line of Table 1: This input contains the verb *think* (subcategorizing for a CP complement) and specifies that its argument has to contain a syntactic variable  $x_j$  which is in the scope of a question operator  $Q_j$ . Such an input has to be realized by a *wh*-question. (See Section 2.2.1 for a short discussion of the problem of input representations.)

Table 1: Constraint profile for direct object extraction (simplified from Legendre et al., 1995, (22a))

$[Q_j [\text{think}_{\text{CP}} [x_j]]]$	SUBCAT	BAR4	BAR3	BAR2	*t
a. $\text{what}_j \text{ do } [\text{you } [\text{think } [\text{he } [\text{said } t_j]]]]$	*		*		*
b. $\text{what}_j \text{ do } [\text{you } [\text{think } [t_j \text{ that } [\text{he } [\text{said } t_j]]]]]$				**	**
c. $\text{what}_j \text{ do } [\text{you } [\text{think } [\text{that } [\text{he } [\text{said } t_j]]]]]$		*			*

Possible realizations of this input are the candidates (a)–(c) in Table 1. These candidates violate different constraints, as indicated by the asterisks in Table 1. For example, candidate (a) violates SUBCAT (as the verb takes an IP complement, instead of a CP complement), \*t (due to the moved *wh*-element it contains), and BAR3 (because the movement crosses three barriers).

The *optimal* structure in a candidate set is computed as the structure that violates the least highly ranked constraints. As an example, consider the competition between candidates (a) and (c): (a) violates SUBCAT, while (c) violates BAR4. According to the constraint hierarchy in (17), SUBCAT is ranked higher than BAR4, which means that candidate (c) wins the competition. Note that all the other constraints that are violated by either of the candidates are not taken into account in determining the winner. Only the most highly ranked constraint on which the two candidates differ matters for the constraint competition (*strict domination* of constraints). Two candidates differ on a constraint if one candidate violates that constraint more often than the other one (e.g., (a) violates SUBCAT once, while (b) violates it zero times).

In Table 1 the optimal candidate is (b): It wins against (c), as it violates BAR2 instead of BAR4. The additional trace that (b) contains allows it to avoid crossing four barriers at once. This means that (b) incurs two violations of \*t (instead of just one). However, this is not relevant to the competition with (c), due to strict domination. (Note that (a) would win if the input contained *think* subcategorizing for an IP.)

Another important aspect of OT can also be illustrated using the extraction example: In OT, crosslinguistic variation can be accounted for by *constraint re-ranking*. Assume that there is an additional constraint \*Q, which disallows empty question operators. For English, the ranking \*Q  $\gg$  \*t holds. This means that questions are formed by movement of *wh*-elements, while in-situ *wh*-elements, which have to be bound by the Q operator, are ungrammatical. Chinese, on the other hand, exhibits the opposite ranking \*t  $\gg$  \*Q, i.e., the use of an empty question operator is preferred to the use of a trace. This explains why in Chinese, *wh*-elements remain in situ in direct object extractions, where the *wh*-element is bound by the Q operator. English, on the other hand, requires *wh*-movement in such configurations, as illustrated by the example in Table 1.

## 2.2. Optimality and Gradient Grammaticality

### 2.2.1. Suboptimal Structures

In line with all major linguistic frameworks, standard OT assumes a binary notion of grammaticality: the competition between candidate structures selects one candidate (or a set of candidates sharing the same constraint profile) as optimal and, hence, grammatical. All losing candidates, i.e., those structures that are *suboptimal*, are assumed to be (equally) ungrammatical; standard OT makes no predictions about the relative ungrammaticality of suboptimal candidates. This binary view of grammaticality is inadequate for data that exhibit a continuum of degrees of acceptability, such as the data reported in the present paper.

We propose to incorporate a gradient notion of grammaticality into OT by dropping the assumption that all suboptimal candidates are (equally) ungrammatical. The aim is to develop an extended version of OT that not only computes the optimal candidate for a given candidate set, but also makes predictions about the relative grammaticality of suboptimal candidates. More specifically, we will assume the following hypothesis:

(18) **Suboptimality Hypothesis**

- a. Suboptimal candidates can differ in grammaticality.
- b. The relative grammaticality of suboptimal candidates can be used as evidence for constraint rankings.

Note that (18b) follows from (18a): if suboptimal candidates differ in grammaticality, then the comparison between two suboptimal candidates can be used as evidence for constraint rankings in the

same way as the comparison between a grammatical candidate and an ungrammatical candidate is used to determine rankings in standard OT.

There are several ways of implementing the suboptimality hypothesis, i.e., of extending OT to make predictions about suboptimal structures (see the discussion in Section 2.2.2 below). A straightforward approach is to generalize the OT notion of grammaticality: in standard OT, grammaticality is defined as global optimality for the whole candidate set. This can be complemented by a definition of *suboptimality* as local optimality relative to a subset of the candidate set. We can then assume that a structure  $S_1$  is less grammatical than a structure  $S_2$  if  $S_1$  is suboptimal with respect to  $S_2$  (see Keller, 1997, for details). Intuitively, this definition entails that the relative grammaticality of a structure corresponds to its *harmony*, i.e., its optimality theoretic rank in the candidate set. This model predicts a grammaticality ordering for the structures in the candidate set, which can then be tested against the acceptability ordering found experimentally for the candidates.

OT employs so-called *ranking arguments* to establish constraint rankings. A ranking argument refers to a set of candidate structures with a certain constraint violation profile, and derives a constraint ranking from this profile. A classical ranking argument works as follows. Assume that two structures  $S_1$  and  $S_2$  that have the same constraint profile, with the following exception:  $S_1$  violates constraint  $C_1$ , but satisfies  $C_2$ . Structure  $S_2$ , on the other hand, violates constraint  $C_2$ , but satisfies  $C_1$ . If  $S_1$  is acceptable but  $S_2$  is unacceptable, then we can conclude that the ranking  $C_2 \gg C_1$  holds.<sup>5</sup>

The suboptimality approach allows a new form of ranking argument based on gradient acceptability data: for suboptimal structures ranking arguments can be based on relative acceptability, instead of binary acceptability, based on Definition (19). (We use the notation  $x > y$  to indicate that the structure  $x$  is more acceptable than the structure  $y$ .)

**(19) Suboptimal Ranking Argument**

- a. The structures  $S_1$  and  $S_2$  form a *minimal pair* if: (i)  $S_1$  violates constraint  $C_1$  and satisfies  $C_2$ ; (ii)  $S_2$  violates constraint  $C_2$  and satisfies  $C_1$ ; and (iii)  $S_1$  and  $S_2$  do not differ on any other constraints.
- b. If  $S_1$  and  $S_2$  form a minimal pair, and  $S_1 > S_2$ , then  $C_2 \gg C_1$ .

Definition (19) instantiates the suboptimality hypothesis in (18b). Note that in a suboptimal ranking argument, both structures in a minimal pair can be suboptimal (i.e., unacceptable to a certain degree), while in a conventional ranking argument, one of the structure has to be optimal. This means that minimal pairs for suboptimal ranking arguments will be much more readily available than minimal pairs for conventional ranking arguments. We will make extensive use of suboptimal ranking arguments in our discussion of the experimental results presented in the remainder of this paper.

A model of grammatical competition has to provide a way of specifying which candidate structures are involved in the competition. In Optimality Theory, this is achieved by specifying an *input*, i.e., a representation from which a set of competing candidate structures are generated. A diverse number of proposals have been put forward in the OT literature regarding which representations are adequate as inputs. Proposals include predicate argument structures (Legendre et al., 1995), sets of lexemes (Grimshaw, 1997), LFG-style f-structures (Bresnan, 2000), or syntactic derivations

<sup>5</sup>This reasoning only covers the simplest type of ranking argument. In the general case, the fact that  $S_1$  is acceptable but  $S_2$  is unacceptable allows the conclusion that each constraint violated by  $S_1$  is outranked by at least one constraint violated by  $S_2$ . See Hayes (1997) for a more extensive discussion of the inference patterns involved in ranking argumentation. In the present paper, we will only deal with simple ranking arguments of the type covered by Definition (19) (which could be generalized straightforwardly).

(Müller, 1999). For the purpose of this paper, minimal assumptions concerning the input are sufficient. We assume that the input specifies the set of constituents (subject, verb, object) that are to be realized by the utterance. The candidate set then contains all permutations of these constituents, possibly marked with accent and augmented with functional elements such as clitics. In addition, the input specifies the Information Structure of the utterance, i.e., each of the constituents is marked as either focus or ground. The candidate set contains only candidates with the same Information Structure (uttered in the same context). As a consequence, candidates differ in how well they map a given linguistic structure on a given Information Structure.

### 2.2.2. *Alternative Approaches*

The suboptimality hypothesis discussed in the previous section allows a flexible treatment of gradient data. There is, however, evidence that not all constraint violations give rise to gradient grammaticality (Keller, 2000). Some constraints seem to induce conventional binary reactions, and a model of gradience has to take this into account. A number of two-level models have been proposed in the literature that accommodate a concept of gradience while preserving a binary notion of grammaticality for certain constraints (Keller, 2000; Müller, 1999). Such a two-level approach is compatible with the analysis of the syntax/phonology interaction in the present paper.

We will illustrate this with respect to Müller's (1999) approach. Müller assumes a distinction between grammaticality (manifested in binary judgments) and markedness (associated with preferences). Grammaticality is handled in terms of standard OT-style constraint competition. All candidates that are suboptimal in this competition are predicted to be categorically ungrammatical. For certain phenomena, the competition will produce not a single optimal candidate, but a set of optimal candidates. All of these candidates are predicted to be grammatical; however, they take part in a further optimality theoretic competition based on a separate set of constraints, so-called markedness constraints. The optimal candidate in this competition is *unmarked*; the suboptimal candidates are more or less marked (dispreferred) depending on their relative suboptimality.

The main attraction of Müller's approach is that it integrates a binary notion of grammaticality (which is clearly adequate for some linguistic phenomena) with a model of relative grammaticality (markedness in his terms).<sup>6</sup> For the analysis of the syntax/phonology interaction in the present paper, we will assume that the competition filtering out candidates that are categorically ungrammatical has already taken place; all the constraints proposed as part of our account are meant to work on the output of this competition. (Examples for constraints that lead to binary ungrammaticality when violated are constraints on agreement or phrase structure.)

It is worth mentioning an alternative proposal for a gradient version of OT that does not rely on an explicit distinction between two types of constraints (Boersma, 1998; Boersma & Hayes, 2001; Hayes, 2000). This approach is based on the assumption that constraint rankings are not fixed, but allow a certain amount of variation. Assume that two candidates  $C_1$  and  $C_2$  are ranked  $C_1 \gg C_2$ , predicting the structure  $S_1$  as optimal for a given input. Under Boersma and Hayes's approach, the assumption is that a re-ranking of  $C_1$  and  $C_2$  can occur with a certain probability, resulting in the ranking  $C_2 \gg C_1$  and producing an alternative optimal candidate  $S_2$ . The degree of grammaticality of  $S_2$  depends on the probability of the re-ranking that makes it optimal. This leads to a continuous notion of grammaticality: the more probable the re-ranking  $C_2 \gg C_1$ , the higher the degree of grammaticality of  $S_2$ ; if the rankings  $C_1 \gg C_2$  and  $C_2 \gg C_1$  are equally probable, then  $S_1$

<sup>6</sup>A similar distinction is expressed by Keller's (2000) dichotomy of hard and soft constraints. The violation of soft constraints leads to gradience, while the violation of hard constraints induces conventional binary judgments.

and  $S_2$  have the same degree of grammaticality. Note that this model makes an (implicit) distinction between binary grammaticality (which obtains if the re-ranking probability of a given constraint is zero), and gradient grammaticality (which emerges from non-zero re-ranking probabilities).

### 2.3. Constraints on Information Structure

Based on the observations outlined in Section 1.2, we propose a set of linguistic constraints that govern the realization of Information Structure in Greek. The purpose of these constraints is to facilitate a systematic discussion of the data and to exemplify how an OT-based approach can capture basic aspects of the experimental results. It is outside the scope of the present paper to provide a full-fledged linguistic analysis of the phenomena under consideration, and we will therefore restrict ourselves to a fairly descriptive formulation of the constraints (for more linguistically sophisticated OT accounts of Information Structure and word order, see Choi, 1996; Müller, 1999; Samek-Lodovici, 1996).

The constraints in (20) are based on our generalizations on Information Structure summarized in (13), and on the observations regarding clitic doubling and verb final orders discussed at the end of Section 1.2.

#### (20) Constraints on Information Structure

- a. GROUNDALIGN (GAGN): ground constituents have to be peripheral.
- b. DOUBLEGROUND (DOUG): clitic doubled objects have to be interpreted as ground.
- c. ACCENTALIGN (ACCAGN): in broad focus, accent has to fall on the rightmost constituent.
- d. ACCENTFOCUS (ACCF): accented constituents have to be interpreted as focus.
- e. DOUBLEALIGN (DOUAGN): preverbal objects have to be clitic doubled.
- f. VERBALIGN (VAGN): the verb must not be right peripheral.

GROUNDALIGN and DOUBLEGROUND impose restrictions on the syntactic/morphological realization of Information Structure. Thus, the association of doubled NPs with a ground interpretation is captured by DOUBLEGROUND. On the other hand, GROUNDALIGN encodes the restriction that ground NPs should appear either to the left or right periphery of the clause. We use the term ‘periphery’ descriptively, to refer to clause initial and clause final NPs. In this respect, our notion of periphery is distinct from the view of periphery taken in syntactic analyses, where occurrence to a peripheral position is associated with a particular syntactic operation (e.g., movement, base-generation, right dislocation). For our purposes, any clause initial or clause final NP is considered peripheral (even if it would correspond to an in-situ constituent in standard syntactic analyses). It is further worth pointing out that this restriction is not biconditional; peripheral NPs do not necessarily belong to the ground part of the sentence. Finally, we interpret GROUNDALIGN as a requirement that *all* ground NPs are peripheral. As we shall see, this point will become relevant in the analysis of V focus, a context that involves two ground NPs.

While GROUNDALIGN and DOUBLEGROUND encode syntactic/morphological restrictions on ground elements, ACCENTFOCUS and ACCENTALIGN are phonological constraints on the realization of focused NPs. ACCENTFOCUS associates an accented constituent with a focus interpretation. It applies to all Information Structures, i.e., both in narrow and broad focus contexts. Moreover, ACCENTFOCUS is insensitive to other structural properties of the relevant constituent (e.g., whether the constituent is an NP or not, whether it appears preverbally or postverbally). In contrast to this,

ACCENTALIGN is only relevant for broad focus, and moreover, it associates accent placement with clause structure (the right clause boundary).

The first four constraints restrict the realization of Information Structure (see (13)), while the last two constraints impose restrictions on word order, independent of information structural factors. DOUBLEALIGN requires preverbal objects to be doubled, while VERBALIGN penalizes verb final orders. (For a more detailed motivation of DOUBLEALIGN, see Section 4.6, where we discuss the results of Experiment 2, which was designed to test this constraint.)

The constraints in (20) conflict in a typical OT fashion. As an example, consider the orders Ovs and Oclvs in an object focus context. Ovs violates DOUBLEALIGN as the object is not clitic doubled; Oclvs, on the other hand, satisfies DOUBLEALIGN. However, this comes at the price of violating DOUBLEGROUND, as the doubled object is not part of ground.

The experiments presented in the remainder of the paper serve a double purpose. On the one hand, the experimental data allow us to test the validity of the set of constraints in (20). The experimental stimuli include structures that violate a given constraint. The constraint violation is expected to lead to a reduction in acceptability compared to structures that fail to incur the violation. The second purpose is to establish a constraint ranking for the constraints in (20). This can be achieved by comparing the degree of unacceptability that ensues from the violations of two distinct constraints, i.e., by using a suboptimal ranking argument as defined in (19).

To preview the main result of Experiments 1 and 2, we give the constraint ranking that will be derived from the experimental data:

$$(21) \{ \text{DOUBLEGROUND, ACCENTFOCUS} \} \gg \{ \text{GROUNDALIGN, VERBALIGN} \} \gg \text{DOUBLEALIGN} \gg \text{ACCENTALIGN}$$

Detailed evidence for this ranking will be discussed in Sections 3.5 and 4.5. Note that the ranking is consistent with the discussion at the end of Section 1.2, where it was argued that accent placement and doubling are the strongest information structural constraints. This was attributed to the unambiguous way in which they mark focused and ground constituents, respectively.

#### 2.4. Magnitude Estimation

The present study relies on very subtle linguistic intuitions, viz., on judgments about the relative acceptability of information structurally different realizations of a sentence. Such intuitions about relative acceptability should be measured experimentally, since the informal elicitation technique traditionally used in linguistics is unlikely to be reliable here (Cowart, 1997; Schütze, 1996; Sorace, 1992). A suitable experimental paradigm is magnitude estimation, a technique standardly applied in psychophysics to measure judgments of sensory stimuli (Stevens, 1975). The magnitude estimation procedure requires subjects to estimate the magnitude of physical stimuli by assigning numerical values proportional to the stimulus magnitude they perceive. Highly reliable judgments can be achieved for a whole range of sensory modalities, such as brightness, loudness, or tactile stimulation.

The magnitude estimation paradigm has been extended successfully to the psychosocial domain (Lodge, 1981) and recently Bard et al. (1996) and Cowart (1997) showed that linguistic judgments can be elicited in the same way as judgments of sensory or social stimuli. In contrast to the 5- or 7-point scale conventionally used to measure human intuitions, magnitude estimation employs a continuous numerical scale. It provides fine-grained measurements of linguistic acceptability, which are robust enough to yield statistically significant results, while being highly replicable



both within and across speakers. Since magnitude estimation provides data on an interval scale, parametric statistics can be used for evaluation.

Magnitude estimation requires subjects to assign numbers to a series of linguistic stimuli proportional to the acceptability they perceive. First, subjects are exposed to a modulus item, to which they assign an arbitrary number. Then, all other stimuli are rated proportional to the modulus, i.e., if a sentence is three times as acceptable as the modulus, it gets three times the modulus number, etc.

### 3. Experiment 1: Interaction of Word Order and Context

#### 3.1. Introduction

Experiment 1 has a triple purpose. Firstly, it investigates the basic claim that word order plays an information structural role in a free word order language like Greek. We elicit acceptability judgments for a variety of word orders and contexts, which allows us to test for an interaction of word order and context. Secondly, the experiment is designed to assess the effect of three constraints: the word order constraint VERBALIGN, the constraint on clitic doubling DOUBLEALIGN, and the constraint GROUNDALIGN regulating the interaction of word order and Information Structure (see (20) for details). The experiment includes sentences that violate one or more of these constraints, and the prediction is that such violations lead to a reduction in acceptability. Thirdly, the experiment tests the suboptimality hypothesis (see (18)). Based on this hypothesis we can derive a constraint hierarchy which predicts word order preferences for a given context. These preferences can then be tested against the acceptability patterns obtained experimentally.

The experimental design includes two factors: word order (ORD) and context (CON). Six word orders were tested: SVO, OVS, VSO, VOS, SOV, and OSV, as illustrated by the following examples:

- (22) a. **SVO:** O Tasos      tha diavasi tin efimerida.  
           the Tasos-NOM will read-3SG the newspaper-ACC  
           ‘Tasos will read the newspaper.’  
       b. **OVS:** Tin efimerida tha diavasi o Tasos.  
       c. **VSO:** Tha diavasi o Tasos tin efimerida.  
       d. **VOS:** Tha diavasi tin efimerida o Tasos.  
       e. **SOV:** O Tasos tin efimerida tha diavasi.  
       f. **OSV:** Tin efimerida o Tasos tha diavasi.

Clitic doubled sentences were not included in this experiment, in order to keep the design at a manageable size. Note that DOUBLEALIGN can be tested on structures that do not contain doubling: for instance, OVS (that violates DOUBLEALIGN) can be compared with SVO (that does not violate DOUBLEALIGN). (Clitic doubled stimuli were included Experiment 2, allowing a direct comparison of OVS with OclVS.)

For the context factor we employed a question context to establish a pattern of ground and focus information, a technique that is widely used in the theoretical literature (e.g., Vallduví, 1992). A total of five contexts were used: null, all focus, subject focus, object focus, and verb focus. As an example, consider the contexts for the sentences in (22):

- (23) a. **Null**  
 b. **All Focus:** Ti tha gini?  
           ‘What will happen?’  
 c. **S Focus:** Pios tha diavasi tin efimerida?  
           ‘Who will read the newspaper?’  
 d. **O Focus:** Ti tha diavasi o Tasos?  
           ‘What will Tasos read?’  
 e. **V Focus:** Ti tha kani o Tasos me tin efimerida?  
           ‘What will Tasos do with the newspaper?’

The null context was included as a control condition, allowing us to study how subjects react in the absence of any contextual information.

### 3.2. Predictions

#### 3.2.1. Null Context Condition

The general prediction is that some word orders are more acceptable than others. Hence we expect to find a main effect of ORD (word order).

Furthermore, the constraints in (20) allow us to make detailed predictions about the acceptability of individual orders. If a given structure violates one of the constraints in (20), then we predict its acceptability to be reduced compared a structure that does not incur this constraint violation. Only the constraints VERBALIGN, DOUBLEALIGN, and GROUNDALIGN are relevant for the present experiment. The other three constraints (DOUBLEGROUND, ACCENTALIGN, and ACCENTFOCUS) deal with clitic doubling and accent placement, and will be investigated in Experiment 2. Note that our predictions only deal with the effect of individual constraint violations; the interaction of constraints (i.e., constraint ranking) will be discussed separately in Section 3.5.

VERBALIGN requires that verbs must not occur at the right periphery of a sentence. This constraint is violated by verb final sentences, where the verb is right peripheral (SOV and OSV sentences in our stimulus set). Hence we expect these orders to be reduced in acceptability. The constraint DOUBLEALIGN requires preverbal objects to be clitic doubled. This constraint is violated by OVS, SOV, and OSV. These orders contain preverbal objects that are not doubled and hence are predicted to be reduced in acceptability. The constraint GROUNDALIGN requires ground constituents to be sentence peripheral. This constraint does not apply in the null context condition, where no information about ground and focus is available.

#### 3.2.2. Context Condition

The general prediction in the context condition is that context has an influence on word order preferences. Hence we expect an interaction of ORD (word order) and CON (context).

Again, the constraints in (20) make predictions based on individual constraint violations. On the one hand, we expect effects from VERBALIGN and DOUBLEALIGN. These are syntactic constraints that are not context dependent. Hence their effects in the context condition should be the same as in the null context condition, i.e., VERBALIGN should disfavor verb final orders (SOV and OSV), while DOUBLEALIGN should disfavor OVS, SOV, and OSV, as these orders include preverbal non-doubled objects.

As for the interaction of word order and context, we expect that the order preferences for each context will reflect the optimal realization of the Information Structure required for this context.

More specifically, the constraint *GROUNDALIGN* predicts that orders with non-peripheral ground constituents will be reduced in acceptability. In the following, we will discuss the predictions for each context. (As in the null context, these predictions only deal with the effect of individual constraint violations; constraint interaction will be discussed separately in Section 3.5.)

*All Focus Context.* There are no ground constituents in the all focus context, hence *GROUNDALIGN* is vacuously satisfied. The order preferences only depend on *VERBALIGN* and *DOUBLEALIGN*. The all focus context is therefore predicted to exhibit the same pattern of word order preferences as the null context.

*S Focus Context.* In the S focus context, the subject is in focus, while the object is part of ground. VOS violates *GROUNDALIGN*, as the object is non-peripheral, and is thus predicted to be less acceptable than SVO, OVS, and VSO, which all satisfy *GROUNDALIGN*. *GROUNDALIGN* is also violated in SOV, which is therefore predicted to be less acceptable than OSV (both orders also violate *VERBALIGN*, and hence should be generally low in acceptability).

*O Focus Context.* In the O focus context, the object is in focus, while the subject is part of ground. This means that *GROUNDALIGN* is violated in VSO, where the subject is non-peripheral. Hence VSO should be dispreferred compared to SVO, OVS, and VOS, which satisfy *GROUNDALIGN*. OSV also incurs a *GROUNDALIGN* violation, and hence should be less acceptable than SOV (both orders also violate *VERBALIGN*).

*V Focus Context.* In the V focus context, the verb is in focus, while both the subject and the object are ground constituents. According to *GROUNDALIGN*, both NPs have to appear in peripheral positions, i.e., clause final or clause initial. It follows that all orders except SVO and OVS violate *GROUNDALIGN*. Thus, VSO, VOS, SOV and OSV are predicted to be reduced in acceptability compared with SVO and OVS. However, as OVS violates *DOUBLEALIGN*, SVO is expected to be the best order. The two final orders, SOV and OSV should be least acceptable: unlike VSO, VOS, and OVS, they violate three constraints (*VERBALIGN*, *DOUBLEALIGN* and *GROUNDALIGN*).

### 3.3. Method

#### 3.3.1. Subjects

Forty native speakers of Greek participated in the experiment. The subjects were recruited over the Internet by postings to newsgroups and mailing lists related to Greece and Greek culture. Participation was voluntary and unpaid. It was made sure that subjects were naive, i.e., they were neither linguists nor students of linguistics.

The data of three subjects were excluded as they were bilingual (by self-assessment). The data of one further subject were excluded as she was a speaker of Cypriot Greek.<sup>7</sup> In addition, the data of two subjects were excluded after an inspection of the responses showed that they had not completed the task adequately.<sup>8</sup> This left 34 subjects for analysis. Of these, 19 subjects were male,

<sup>7</sup>Cypriot Greek is a dialect that differs considerably from standard Greek. It is not clear whether the differences between Cypriot and standard Greek would affect the current study, but for methodological reasons, it was decided to exclude speakers of Cypriot Greek.

<sup>8</sup>Subjects were excluded based on response times and response ranges. Appendix A contains a more detailed description of the data recorded by the experimental software this purpose.

15 female; five subjects were left-handed, 29 right-handed. The age of the subjects ranged from 21 to 42 years, the mean was 26.7 years.

### 3.3.2. *Materials*

*Training Materials.* The experiment included a set of training materials that were designed to familiarize subjects with the magnitude estimation task. The training set contained six horizontal lines. The range of largest to smallest item was 1:6.7. The items were distributed evenly over this range, with the largest item covering the maximal window width of the web browser. A modulus item in the middle of the range was provided.

*Practice Materials.* A set of practice items was used to familiarize subjects with applying magnitude estimation to linguistic stimuli. The practice set consisted of six sentences that were representative of the test materials. A wide spectrum of acceptability was covered, ranging from fully acceptable to severely unacceptable. A modulus item in the middle of the range was provided.

*Test Materials.* For the experimental items, a full factorial design was used with word order (ORD) and context (CON) as the two factors (see (22) and (23) for example stimuli). This yielded a total of  $ORD \times CON = 6 \times 5 = 30$  cells. Eight lexicalizations per cell were used, which resulted in a total of 240 stimuli.

A set of 24 fillers was used, designed to cover the whole acceptability range. Six items of each of the following four groups were used: no violation, case violation, phrase structure violation, and agreement violation. The fillers covered a range of word orders, including ones that were not used in the experimental items (e.g., by using null subjects). The contexts for the fillers included *wh*-questions (both adjunct and complement questions) and *yes-no*-questions. As in the practice phase, a modulus item in the middle of the range was provided (see Appendix B for a list of all experimental materials).

### 3.3.3. *Procedure*

The method used was magnitude estimation as proposed by Lodge (1981) and extended to linguistic stimuli by Bard et al. (1996). Each subject took part in an experimental session that lasted approximately 15 minutes and consisted of a training phase, a practice phase, and an experimental phase. The experiment was self-paced, though response times were recorded to allow the data to be screened for anomalies.

The experiment was conducted remotely over the Internet. The subject accessed the experiment using his or her web browser. The browser established an Internet connection to the experimental server, which was running WebExp 2.1 (Keller, Corley, Corley, Konieczny, & Todirascu, 1998), an interactive software package for administering web-based psychological experiments. (For a detailed discussion of the validity of web-based data, see Appendix A.)

*Instructions.* Before the actual experiment started, a set of instructions in Greek was presented. The instructions first explained the concept of numerical magnitude estimation of line length. Subjects were instructed to make estimates of line length relative to the first line they would see, the reference line. Subjects were told to give the reference line an arbitrary number, and then assign a number to each following line so that it represented how long the line was in proportion to the reference line. Several example lines and corresponding numerical estimates were provided to illustrate the concept of proportionality.

Then subjects were told that linguistic acceptability could be judged in the same way as line length. The concept of linguistic acceptability was not defined; instead, examples of acceptable and unacceptable sentences were provided, together with examples of numerical estimates.

Subjects were told that they could use any range of positive numbers for their judgments, including decimals. It was stressed that there was no upper or lower limit to the numbers that could be used (exceptions being zero or negative numbers). Subjects were urged to use a wide range of numbers and to distinguish as many degrees of acceptability as possible. It was also emphasized that there were no ‘correct’ answers, and that subjects should base their judgments on first impressions, not spending too much time to think about any one sentence.

*Demographic Questionnaire.* After the instructions, a short demographic questionnaire was administered. The questionnaire included name, email address, age, sex, handedness, academic subject or occupation, and language region. Handedness was defined as ‘the hand you prefer to use for writing’, while language region was defined as ‘the place (city, region/state/province, country) where you learned your first language’. The results of the questionnaire were reported in the Subjects section above.

*Training Phase.* The training phase was meant to familiarize subjects with the concept of numerical magnitude estimation using line lengths. Items were presented as horizontal lines, centered in the window of the subject’s web browsers. After viewing an item, the subject had to provide a numerical judgment over the computer keyboard. After pressing Return, the current item disappeared and the next item was displayed. There was no possibility to revisit previous items or change responses once Return had been pressed. No time limit was set for either the item presentation or for the response.

Subjects first judged the modulus item, and then all the items in the training set. The modulus remained on the screen all the time to facilitate comparison. Items were presented in random order, with a new randomization being generated for each subject.

*Practice Phase.* This phase allowed subjects to practice magnitude estimation of linguistic acceptability. Presentation and response procedure was the same in the training phase, with linguistic stimuli being displayed instead of lines. Each subject judged the whole set of practice items.

As in the training phase, subjects first judged the modulus item, and then all the items in the training set. The modulus remained on the screen all the time to facilitate comparison. Items were presented in random order, with a new randomization being generated for each subject.

*Experimental Phase.* Presentation and response procedure in the experimental phase was the same as in the practice phase. A between-subjects design was used to administer the experimental stimuli: subjects in Group A judged non-contextualized stimuli, while subjects in Group B judged contextualized stimuli.

For Group A, two test sets were used: each set contained four lexicalizations for each of the six levels of factor ORD, i.e., a total of 24 items. Lexicalizations were assigned to test sets using a Latin square design. For Group B, eight test sets were used: each set contained one lexicalization for each of the six orders in each of the four contexts, i.e., a total of 24 items. Again, a Latin square was used to assign lexicalizations to test sets.

As in the practice phase, subjects first judged the modulus item, which remained on the screen all the time. Then they saw 48 test items: 24 experimental items and 24 fillers. Items were presented

in random order, with a new randomization being generated for each subject. Each experimental subject was randomly assigned to a group and a lexicalization: 17 subjects were assigned to each group. Instructions, examples, training items, and fillers were adapted for Group B to take context into account.

### 3.4. Results

The data were normalized by dividing each numerical judgment by the modulus value that the subject had assigned to the reference sentence. This operation creates a common scale for all subjects. Then the data were transformed by taking the decadic logarithm. This transformation ensures that the judgments are normally distributed and is standard practice for magnitude estimation data (Bard et al., 1996; Lodge, 1981). All analyses were conducted on the normalized, log-transformed judgments.

All the figures in this paper display means of normalized, log-transformed judgments, together with standard errors. Appendix C contains the descriptive statistics for the experimental results.

#### 3.4.1. Null Context Condition

Separate ANOVAs were performed for the context and the null context condition. The mean judgments for the null context condition are graphed in Figure 1. An ANOVA revealed a significant main effect of word order ( $F_1(5, 80) = 20.00, p < .0005; F_2(5, 35) = 3.18, p = .018$ ). This confirms our general prediction that some word orders are more acceptable than others, even in absence of context.

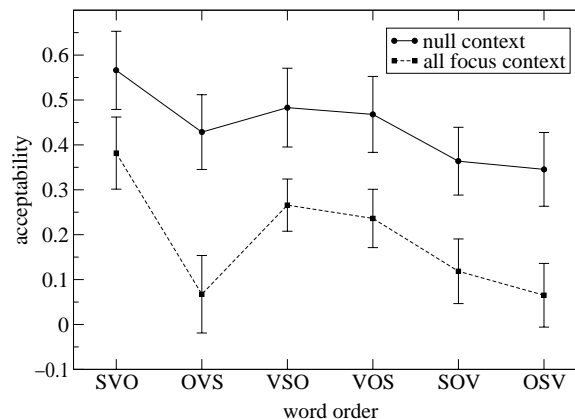


Figure 1. Word Order Preferences in the Null Context and the All Focus Context (Experiment 1)

A post-hoc Tukey test was carried out for the main effect of ORD. This test determines which word orders differ in acceptability and thus allows us to assess the influence of the context independent constraints VERBALIGN and DOUBLEALIGN.

VERBALIGN requires that verbs must not occur clause finally, thus predicting reduced acceptability for the verb final orders SOV and OSV. This was confirmed by the Tukey test, which showed that SOV was significantly less acceptable than SVO (by subjects,  $\alpha < .01$ , and by items,

$\alpha < .05$ ), VSO (by subjects only,  $\alpha < .01$ ), and VOS (by subjects only,  $\alpha < .01$ ). OSV was significantly less acceptable than SVO (by subjects,  $\alpha < .01$ , and by items,  $\alpha < .05$ ), VSO (by subjects only,  $\alpha < .01$ ), and VOS (by subjects only,  $\alpha < .01$ ).

The constraint DOUBLEALIGN requires preverbal objects to be clitic doubled, which means that OVS should be reduced in acceptability compared to SVO. This prediction was confirmed by the Tukey test, which showed that OVS was less acceptable than SVO (by subjects only,  $\alpha < .01$ ). Furthermore, we found that OSV was less acceptable than OVS (by subjects only,  $\alpha < .05$ ). Both orders violate DOUBLEALIGN, but OSV is verb final and hence also violates VERBALIGN, which explains the difference in acceptability.

In addition, we found that SVO was more acceptable than the verb initial orders VSO (by subjects only,  $\alpha < .01$ ) and VOS (by subjects only,  $\alpha < .01$ ). This is unexpected, as neither of these three orders violates any constraints, and we would expect them to be equally acceptable. All other differences failed to reach significance.

#### 3.4.2. Context Condition

The mean judgments for the context condition are graphed in Figures 2–4. As in the null context condition, an ANOVA revealed a significant main effect of word order ( $F_1(5, 80) = 24.97$ ,  $p < .0005$ ;  $F_2(5, 35) = 11.15$ ,  $p < .0005$ ). A marginally significant main effect of context was also found ( $F_1(3, 48) = 2.58$ ,  $p = .064$ ;  $F_2(3, 21) = 3.28$ ,  $p = .041$ ). The interaction of word order and context was also significant ( $F_1(15, 240) = 2.47$ ,  $p = .002$ ;  $F_2(15, 105) = 1.97$ ,  $p = .024$ ), which confirms our general prediction that context has an influence on word order preferences.

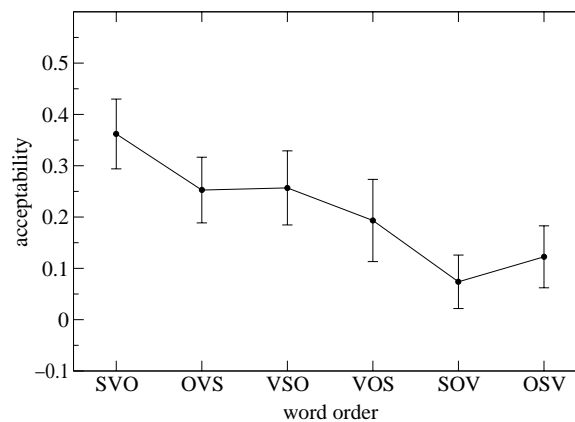


Figure 2. Word Order Preferences in the S Focus Context (Experiment 1)

A post-hoc Tukey test was carried out on the ORD effect to determine the effects of the context independent constraints VERBALIGN and DOUBLEALIGN. The resulting pattern closely matched the one found in the null context condition. Verb final orders were reduced in acceptability, in line with the predictions of VERBALIGN. SOV was significantly less acceptable than SVO ( $\alpha < .01$ ), VSO (by subjects only,  $\alpha < .05$ ), and VOS (by subjects only,  $\alpha < .01$ ). Furthermore, OSV was significantly less acceptable than SVO ( $\alpha < .01$ ), VSO (by subjects only,  $\alpha < .01$ ), and VOS (by subjects,  $\alpha < .01$ , and by items,  $\alpha < .05$ ). We also found that OVS was less acceptable than SVO ( $\alpha < .01$ ), in line with the predictions of DOUBLEALIGN.

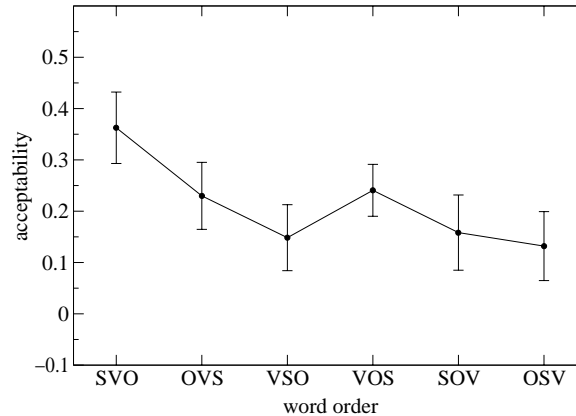


Figure 3. Word Order Preferences in the O Focus Context (Experiment 1)

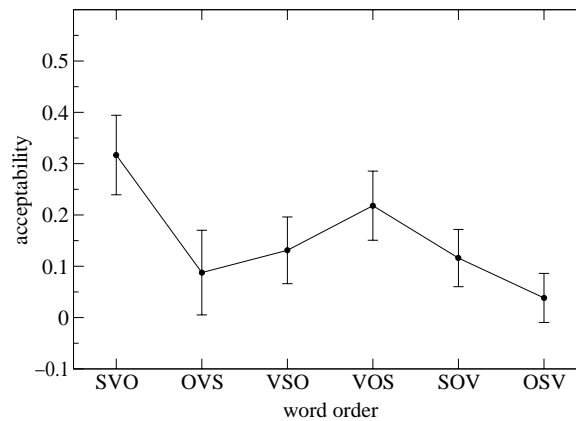


Figure 4. Word Order Preferences in the V Focus Context (Experiment 1)

As in the null context condition, SVO was more acceptable than VSO ( $\alpha < .01$ ) and VOS (by subjects,  $\alpha < .01$ , and by items,  $\alpha < .05$ ). There were no other significant differences.

A further Tukey test was carried for the ORD/CON interaction to assess the effect of the constraint GROUNDALIGN, which predicts that orders with non-peripheral ground constituents will be reduced in acceptability. We will discuss each context separately.

*All Focus Context.* GROUNDALIGN is vacuously satisfied in an all focus context. Hence we predicted that the all focus context will show the same pattern of order preferences as the null context. This prediction was borne out, as illustrated by Figure 1, which compares the two preference patterns.

*S Focus Context.* Here we predicted that VOS, which violates GROUNDALIGN, should be reduced in acceptability compared to SVO, OVS, and VSO, which all satisfy GROUNDALIGN. The Tukey test (see Figure 2) provided a partial confirmation: VOS was significantly less acceptable than



SVO (by subjects only,  $\alpha < .05$ ). However, the differences between VOS and OVS and between VOS and VSO failed to reach significance.

We also predicted OSV to be preferred over SOV, which violates *GROUNDALIGN*. Again, this preference was too small to reach significance. On the other hand, SOV was significantly less acceptable than SVO ( $\alpha < .01$ ), OVS (by subjects only,  $\alpha < .05$ ), and VSO (by subjects only,  $\alpha < .05$ ). OSV was less acceptable than SVO ( $\alpha < .01$ ). These differences are readily explained by the constraint *VERBALIGN*, which is violated in verb final orders, but not in verb initial and verb medial ones. All other differences were not significant.

*O Focus Context.* Here we predicted VSO (violating *GROUNDALIGN*) to be less acceptable than SVO, OVS, and VOS (all satisfying *GROUNDALIGN*). This was partially borne out by the Tukey test (see Figure 3) which demonstrated that VSO was significantly less acceptable than SVO (by subjects,  $\alpha < .01$ , and by items,  $\alpha < .05$ ). However, we failed to find significant differences between VSO and OVS and between VSO and VOS.

We also predicted SOV to be preferred over OSV, which violates *GROUNDALIGN*. Again, this preference was too small to reach significance. On the other hand, we found that the preference SVO > SOV was significant (by subjects,  $\alpha < .01$ , and by items,  $\alpha < .05$ ), as well as the preference SVO > OSV ( $\alpha < .01$ ). This is explained by the fact that the verb final orders violate *VERBALIGN*. There were no other significant differences.

*V Focus Context.* In the V focus context, VSO and VOS violate *GROUNDALIGN* and hence are predicted to be reduced in acceptability compared to SVO, which satisfies *GROUNDALIGN*. The Tukey test (see Figure 4) confirmed this by showing that SVO was significantly more acceptable than VSO ( $\alpha < .05$ ). The difference between SVO and VOS, however, failed to reach significance. On the other hand, OVS was less acceptable than SVO, readily explained by the fact that OVS violates *DOUBLEALIGN*. Also, the preference SVO > SOV was significant (by subjects,  $\alpha < .01$ , and by items,  $\alpha < .05$ ), as well as the preference SVO > OSV ( $\alpha < .01$ ). The low acceptability of the two final orders was expected, as they violate three constraints (*VERBALIGN*, *DOUBLEALIGN*, and *GROUNDALIGN*). There were no other significant differences.

### 3.5. Optimality Theoretic Modeling

In this section we offer an analysis of the experimental results within the OT framework we introduced in Section 2. We first derive a constraint hierarchy, and then turn to comparing the experimental results with the predictions of our model.

The experimental findings provide clear evidence for the validity of the three constraints *VERBALIGN*, *DOUBLEALIGN*, and *GROUNDALIGN* (see (20) for details). *VERBALIGN* predicts that verb final orders are reduced in acceptability. This was confirmed by the fact that SOV and OSV were found to be consistently dispreferred, both in the null context condition and in the context condition. *DOUBLEALIGN* penalizes non-clitic doubled preverbal objects, and is violated by OVS, SOV, and OSV. These orders were clearly dispreferred in both the null context and the context condition. *GROUNDALIGN* requires ground constituents to be sentence peripheral. The effect of this constraint was evident in various contexts. In S focus context, SVO was more acceptable than VOS, which violates *GROUNDALIGN*. In the O focus and V focus context, VSO violates *GROUNDALIGN* and was less acceptable than SVO (see Figures 1–4 for details).

Apart from confirming the validity of our constraint set, the experimental data also allow us to establish a constraint ranking for the three constraints under investigation. This can be achieved

using suboptimal ranking arguments, as defined in (19). To rank three constraints, two ranking arguments are required, which are detailed below. One-way ANOVAs (planned comparisons) are used to determine if the acceptability difference for a given ranking argument is significant. We adjust the significance level using the Bonferroni method, i.e., a significance threshold of  $p = .025$  will be assumed, since two comparisons are carried out.

The stimulus set provides two minimal pairs that can serve as evidence for the ranking of *GROUNDALIGN* with respect to *DOUBLEALIGN*. In the *S* focus context, *OVS* violates *DOUBLEALIGN*, but satisfies *GROUNDALIGN*. *VOS*, on the other hand, satisfies *DOUBLEALIGN*, but violates *GROUNDALIGN*. The two structures differ in no other constraints, and therefore constitute a minimal pair; we observe the preference  $OVS > VOS$ . In the *O* focus context, *OVS* and *VSO* form a minimal pair; the preference is  $OVS > VSO$ . Hence both minimal pairs indicate that  $GROUNDALIGN \gg DOUBLEALIGN$  is the correct ranking. A one-way ANOVA on the combined data from both minimal pairs confirms this: orders that violate *GROUNDALIGN* are significantly less acceptable than orders that violate *DOUBLEALIGN* (by subjects only,  $F_1(1, 16) = 7.90$ ,  $p = .012$ ;  $F_2(1, 7) = 1.95$ ,  $p = .206$ ).

As a next step, we will derive the ranking of *VERBALIGN* with respect to *GROUNDALIGN*. Unfortunately, the data fail to provide a suitable minimal pair, hence we cannot establish a suboptimal ranking argument. However, we can use the following *reductio ad absurdum* argument to arrive at a ranking. In the *S* focus context, *VOS* violates *GROUNDALIGN*, while *OSV* violates *VERBALIGN* and *DOUBLEALIGN*. Under the ranking  $GROUNDALIGN \gg VERBALIGN$  we predict the preference  $OSV > VOS$  (see Figure 2). The same reasoning applies to the *O* focus context, where *VSO* violates *GROUNDALIGN*, while *SOV* violates *VERBALIGN* and *DOUBLEALIGN*. Under the ranking  $GROUNDALIGN \gg VERBALIGN$  we predict  $SOV > VSO$  (see Figure 3). However, a one-way ANOVA on the combined data fail to find a significant acceptability difference between the orders that violate *GROUNDALIGN* and the orders that violate *VERBALIGN* and *DOUBLEALIGN*.

This leads to the conclusion that  $GROUNDALIGN \gg VERBALIGN$  cannot be the correct ranking; either *GROUNDALIGN* and *VERBALIGN* are ranked equally, or  $VERBALIGN \gg GROUNDALIGN$  holds. As we do not have evidence for the latter, we will assume that *GROUNDALIGN* and *VERBALIGN* are ranked equally. By transitivity, it follows that the ranking  $VERBALIGN \gg DOUBLEALIGN$  must hold, leading to the following overall constraint hierarchy:

$$(24) \quad \{VERBALIGN, GROUNDALIGN\} \gg DOUBLEALIGN$$

Based on this hierarchy, we can now compute constraint violation profiles for the structures investigated in Experiment 1. Recall that we made the assumption that the input for a constraint competition specifies the constituents to be expressed, plus the Information Structure of the utterance (see the Section 2.2.1). This means that a separate constraint competition takes place for each context. The constraint profiles for the four contexts investigated in Experiment 1 are given in Tables 2–5.

Under the suboptimality hypothesis in (18), a constraint profile predicts not only the optimal structure for a given candidate set, but also induces a grammaticality order on the candidates, based on their relative suboptimality. This order can then be compared with the acceptability order that was determined experimentally for a given context. We determined the acceptability order by converting the average judgments to ranks, ignoring differences that are smaller than one standard error.<sup>9</sup> Correlation analysis can be used to test the match between the acceptability order and the

<sup>9</sup>More precisely, the following criterion was adopted: two means  $m_1$  and  $m_2$  were considered the same if  $m_2$  was

grammaticality order. We compared the ranks of the candidates using Spearman’s  $\rho$ , which is the appropriate correlation coefficient for ordinal data (the results reported are for a one-tailed test, and have been corrected for tied ranks). The degree of correlation between the theoretical and the experimental orders indicates how well our OT model fits the experimental data.

Table 2: Constraint Profile for the Null Context and the All Focus Context (Experiment 1)

all focus	VAGN	GAGN	DOUAGN
SVO			
OVS			*
VSO			
VOS			
SOV	*		*
OSV	*		*

The profile in Table 2 predicts the following grammaticality order for the null context and the all focus context:  $\{SVO, VSO, VOS\} > OVS > \{SOV, OSV\}$ . In Experiment 1, we found the following order for the null context (see Figure 1):  $SVO > \{VSO, VOS\} > OVS > \{SOV, OSV\}$ . A high correlation between the predicted order and the actual order was obtained ( $\rho(6) = .95$ ,  $p = .002$ ). The main mismatch is due to the fact that VSO and VOS were less acceptable than SVO, which is not predicted by the violation profile in Table 2.

The predicted order for the all focus context was the same as in the null context (see Figure 1), while the experimentally found order was:  $SVO > \{VSO, VOS\} > SOV > \{OVS, OSV\}$ . Again, a high correlation between the predicted and the actual order was obtained ( $\rho(6) = .81$ ,  $p = .025$ ). Note, however, that VSO and VOS were less acceptable than predicted (as in the null context). In addition, OVS was less acceptable than predicted (and than in the null context).

We also found a high correlation between the acceptability order for the all focus context and the null context ( $\rho(6) = .86$ ,  $p = .013$ ). This confirms our hypothesis that even when faced with isolated sentences, native speakers make implicit assumptions about Information Structure—they assume an all focus context.

Table 3: Constraint Profile for the S Focus Context (Experiment 1)

S focus	VAGN	GAGN	DOUAGN
SVO			
OVS			*
VSO			
VOS		*	
SOV	*	*	*
OSV	*		*

For the S focus context, we predict the following order (see Table 3):  $\{SVO, VSO\} > OVS >$  within the interval  $m_1 \pm e_1$ , or if  $m_1$  was within the interval  $m_2 \pm e_2$ , where  $e_1$  and  $e_2$  are the standard errors associated with  $m_1$  and  $m_2$ , respectively. The means and standard errors for all conditions are listed in Appendix C.

VOS > OSV > SOV. The experimental results confirm this acceptability pattern (see Figure 2): SVO > {VSO, OVS} > VOS > OSV > SOV. Again, a high correlation between the predicted order and the actual order was obtained ( $\rho(6) = .95$ ,  $p = .001$ ). The only discrepancy concerns VSO, which the model predicts to be as acceptable as SVO; in the experimental data, however, VSO is equal in acceptability to OVS.

Table 4: Constraint Profile for the O Focus Context (Experiment 1)

O focus	VAGN	GAGN	DOUAGN
SVO			
OVS			*
VSO		*	
VOS			
SOV	*		*
OSV	*	*	*

For the O focus context, our OT model makes the following prediction (see Table 4): {SVO, VOS} > OVS > VSO > SOV > OSV. This is matched well by the experimentally obtained order (see Figure 3): SVO > {VOS, OVS} > {VSO, SOV, OSV}; the correlation between the two orders is high ( $\rho(6) = .89$ ,  $p = .008$ ). The same mismatch occurs as in the S focus context: in the experimental data, the acceptability of VOS is reduced compared to SVO, while the model predicts both orders to be equally acceptable. In addition, the two verb final orders SOV and OVS are more acceptable than expected, patterning with VSO, which violates GROUNDALIGN.

Table 5: Constraint Profile for the V Focus Context (Experiment 1)

V focus	VAGN	GAGN	DOUAGN
SVO			
OVS			*
VSO		*	
VOS		*	
SOV	*	*	*
OSV	*	*	*

Table 5 predicts the following order preferences for the V focus context: SVO > OVS > {VSO, VOS} > {SOV, OSV}. Experimentally we found (see Figure 4): SVO > VOS > {OVS, VSO, SOV} > OSV. The correlation between the predicted order and the experimental order was lower than for the other contexts, and just reached significance ( $\rho(6) = .72$ ,  $p = .054$ ). Note that the acceptability of VOS and SOV was higher than predicted by the model.

To summarize, our OT model of the interaction of VERBALIGN, DOUBLEALIGN, and GROUNDALIGN is generally well supported by the experimental data. However, we observed some mismatches between the experimental data and the prediction of the model. These mismatches are largely related to two problems: there is an unexpected reduction in the acceptability of verb initial orders compared to SVO, and there is an unexpected context effect for the constraint

DOUBLEALIGN. Both issues will be discussed in the next section.

### 3.6. Discussion

The results of the present experiment provide support on three levels for the model of word order preferences put forward in this paper. Firstly, the experimental findings confirm the general hypothesis that Information Structure has an influence on word order preferences in Greek. This was evidenced by the significant interaction of word order and context found in the context condition.

Secondly, the results provide evidence for the constraints VERBALIGN, DOUBLEALIGN, and GROUNDALIGN, which are part of our account of the interaction of syntax, phonology, and Information Structure (see (20) for the full constraint set). We were able to confirm the detailed predictions of the context independent constraints VERBALIGN and DOUBLEALIGN by carrying out a set of post-hoc tests on the word order preferences obtained in both the null context condition and the context condition. The predictions of the context specific constraint GROUNDALIGN were validated via a separate set of post-hoc tests on the interaction of word order and context.

Thirdly, the present experiment provides evidence for the suboptimality hypothesis in (18) by demonstrating that an account of the interaction of word order and context can be developed based on this hypothesis. Using suboptimal ranking arguments (see (19)), the experimental data allow us to derive a hierarchy for the constraints VERBALIGN, DOUBLEALIGN, and GROUNDALIGN (see (24)). It was demonstrated that this hierarchy not only predicts the optimal candidate for a given input, but also induces a grammaticality order on the candidate set, which can then be tested against experimentally elicited preference data. In all five contexts, we achieved a high correlation between the grammaticality order predicted by our model and the acceptability order induced by the experimental data.

There are a number of issues that our model left unresolved. One is the status of DOUBLEALIGN. This constraint seems to be context dependent; it causes strong acceptability effects in the all focus context and the V context, but leads to only small acceptability differences in the null context, S focus context, and O focus context. This is one of the reasons for the mismatches between the order preferences predicted by our model and the ones observed experimentally (in the all focus and V context). We conclude that further experimental studies are necessary to clarify the status of DOUBLEALIGN. No clitic doubled stimuli were included in the present experiment, which makes it hard to assess the effect of DOUBLEALIGN: we cannot check if the clitic doubled version of OVS is really as acceptable as SVO (which is what DOUBLEALIGN predicts). Perhaps OVS is inherently less acceptable than SVO, even under clitic doubling. We will return to this point in Experiment 2, which includes clitic doubled stimuli.

Another issue concerns the status of the verb initial orders VSO and VOS. The experimental data show that the acceptability of these orders is generally reduced compared to SVO. This holds even when the verb initial orders incur no constraint violations and thus are predicted to be as acceptable as SVO. This is an unexpected result in view of the set of constraints in (20), and it is unclear how this finding can be explained. However, it seems unlikely that an explanation in terms of Information Structure is possible. As will be shown below, the effect disappears in Experiment 2. This might be due to the fact that Experiment 2 used speech stimuli, while Experiment 1 was based on written stimuli. As the written language is typically associated with a more formal register, it seems plausible to assume that written stimuli trigger a more normative behavior in the subjects. This would explain the preference for SVO over verb initial orders, as SVO is typically assumed to be the ‘correct’ word order in prescriptive grammars of Greek.

Another result of Experiment 1 concerns the null context condition: here, we found the same pattern as in the all focus context (see Figure 1). This is an important methodological finding, as it indicates that even when faced with isolated sentences (which have traditionally been the focus of syntactic research), native speakers make implicit assumptions about Information Structure—they assume an all focus context. We will include the same null context condition in Experiment 2 to test the generality of this result.

Figure 1 also illustrates another interesting effect. The overall acceptability in the all focus context is consistently lower than in the null context, even though the acceptability pattern is the same in both contexts.<sup>10</sup> This observation also holds for the other context conditions, which receive lower overall acceptability judgments than the null context (see Figures 2–4). A possible explanation for this effect is that the present experiment used relatively artificial question-answer pairs in which the answer repeats material from the question (see (22) and (23)). A more natural answer would contain a pronoun instead of repeating a ground NP. Also, it would be natural to drop ground subjects in the answer (Greek is a pro-drop language). While the stimuli in the all focus context do not contain repeated material, they still provide rather artificial dialogues: *Peter will read the newspaper* is an unexpected answer to a question like *What's new?*. Note, however, that consistent information structural effects can be observed even if artificial contexts are used, as the results of the present experiment have demonstrated.

It is important to keep in mind that the present paper is not concerned with absolute acceptability, but rather deals with relative acceptability, i.e., with acceptability differences triggered by constraint violations. There are numerous non-linguistic factors that influence the absolute value of acceptability judgments, including type of instructions, type of fillers used, and the modality of the stimuli (spoken or written) (see Schütze, 1996, and Cowart, 1997, for surveys). This means that the absolute acceptability of a given sentence is expected to vary from experiment to experiment, while the relative acceptability of two stimuli can be expected to be constant across experiments.

#### 4. Experiment 2: Interaction of Word Order, Accent Placement, Clitic Doubling, and Context

##### 4.1. Introduction

Experiment 2 is designed to answer three main questions. Firstly, it investigates the basic claim that clitic doubling and accent placement play an information structural role in a free word order language like Greek. Secondly, the experiment extends the results of Experiment 1 by investigating the validity of a total of five constraints: the word order constraints GROUNDALIGN, the clitic doubling constraints DOUBLEALIGN and DOUBLEGROUND, and accent constraints ACCENTALIGN and ACCENTFOCUS (see (20) for details). Thirdly, the experimental data will provide further evidence for the hierarchical nature of these constraints thus providing additional support for the suboptimality hypothesis (see (18)).

Experiment 2 employs a full factorial design involving the following factors: word order (ORD), clitic doubling (DOU), accent placement (ACC), and context (CON). In order to keep the design at a manageable size, only three word orders were included: SVO, OVS, VSO. The order VOS behaved essentially symmetric to VSO in Experiment 1, and was therefore excluded from the present design. The verb final orders were also excluded, as they were mainly used to establish the validity of VERBALIGN, and hence are not essential for the present experiment.

<sup>10</sup>We owe this observation to Mark Steedman and an anonymous reviewer.

The factor DOU had two levels: clitic doubled object and non-doubled object. The following examples represent the clitic doubled versions of the example stimuli in (22):

- (25) a. **ScIVO:** O Tasos tha tin diavasi tin efimerida.  
           the Tasos-NOM will it-CL read-3SG the newspaper-ACC  
           ‘Tasos will read the newspaper.’  
       b. **OcIVS:** Tin efimerida tha tin diavasi o Tasos.  
       c. **clVSO:** Tha tin diavasi o Tasos tin efimerida.

The accent factor ACC also had two levels: accent on the subject, and accent on the object; consider the following examples:

- (26) a. **Svo:** O TASOS tha diavasi tin efimerida.  
       b. **svO:** O Tasos tha diavasi tin EFIMERIDA.

We used the same four contexts for factor CON as in Experiment 1, illustrated in (23). Again a null context was included as a control condition, enabling us to test the hypothesis that isolated sentences are judged like sentences in an all focus context.

To limit the complexity of the experimental design, we did not include a V accent condition. This means that there is no appropriate intonational realization for the V focus context, where accent is preferred on V. However, we still expect the preference profile for V focus to be informative, as it allows us to investigate the behavior of suboptimal accent realizations (S accent and O accent). Furthermore, the V focus condition is necessary for a full comparison of the results of Experiment 2 with the context effects found in Experiment 1.

## 4.2. Predictions

### 4.2.1. Null Context Condition

A general prediction is that the acceptability of certain orders (such as OVS) will be affected by clitic doubling. Hence an interaction of ORD and DOU (clitic doubling) should be present. An interaction of ORD and ACC (accent placement) is also expected: sentence final accent is preferred in the null context (assuming that a null context behaves like an all focus context), hence some orders will prefer subject accent, while others will prefer object accent. Finally, we predict an interaction of DOU and ACC. This follows from the unacceptability of accented clitic doubled objects (see Section 1.2 for details). As in the Experiment 1, the present predictions only deal with the effect of individual constraint violations; the interaction of constraints (i.e., constraint ranking) will be discussed separately in Section 4.5.

Furthermore, the constraints in (20) allow us to make detailed predictions about the acceptability of individual orders. If a given structure violates one of the constraints in (20), then we predict its acceptability to be reduced compared a structure that does not incur this constraint violation. These predictions can be tested by further investigating the main effect of ORD and the pairwise interactions of ORD, DOU, and ACC. Table 6 details which effects will be used to test which constraints. Note that the VERBALIGN, requiring verbs not to be right peripheral, is not relevant, as no verb final orders were included in the present experiment. DOUBLEALIGN, which states that preverbal objects have to be clitic doubled, is violated by OVS. OVS is therefore predicted to be dispreferred compared to SVO and VSO. However, the difference between OVS and SVO/VSO should disappear in clitic doubled orders, where OcIVS satisfies DOUBLEALIGN.

Table 6: Main Effects and Interactions Used to Test the Constraint Set (Experiment 2, Null Context Condition)

Interaction	Constraints
ORD	GROUNDALIGN
ORD/DOU	DOUBLEALIGN
DOU/ACC	ACCENTFOCUS, DOUBLEGROUND
ORD/ACC	ACCENTALIGN

Experiment 1 provided evidence for the hypothesis that a null context behaves like an all focus context. Under this assumption, we can derive predictions from the information structural constraints GROUNDALIGN, DOUBLEGROUND, ACCENTFOCUS, and ACCENTALIGN by treating the null context as an all focus context. The constraint GROUNDALIGN, which states that ground constituents have to be sentence peripheral, is vacuously satisfied—there are no ground constituents in an all focus context. The same holds for ACCENTFOCUS, which requires accented constituents to be interpreted as focus. All constituents are in focus, i.e., this constraint is always satisfied, no matter what the accent pattern is.

An interesting case is DOUBLEGROUND, which states that clitic doubled objects have to be interpreted as ground. In stimuli with clitic doubling, DOUBLEGROUND imposes an interpretation in which the object is ground. However, as discussed in Section 1.2, an all focus context may accept a wider range of felicitous answers, including answers with doubled objects (see examples (11) and (12)). Hence we do not expect an effect of DOUBLEGROUND here. We do, however, predict reduced acceptability for stimuli with accented doubled objects: DOUBLEGROUND states that doubled objects are interpreted as ground; ACCENTFOCUS, however, requires accented constituents to be interpreted as focus. This leads to an inherent, context independent constraint conflict in orders with object accent and clitic doubling, which are, therefore, predicted to be dispreferred over clitic doubled orders with subject accent and all non-clitic doubled orders.

Finally, ACCENTALIGN requires that in broad focus, accented constituents have to be right peripheral. Since the null focus context behaves like an all focus context (i.e., a broad focus context), orders with clause final accent are expected to be preferred: thus, svO should be preferred over Svo, ovS over Ovs, and vsO over vSo. Similarly, for stimuli involving clitics, ACCENTALIGN predicts that ScvO, Oclvs, and clvSo will be reduced in acceptability.

#### 4.2.2. Context Condition

On a general level, we expect to find the effects involving ORD, DOU, and ACC that were predicted for the null context condition, i.e., we expect the interactions ORD/DOU, DOU/ACC, and ORD/ACC. The second general prediction is that the accent placement and clitic doubling will interact with Information Structure. Hence, we expect interactions of ACC and CON and of DOU and CON. In addition, the interaction of ORD and CON that was detected in Experiment 1 should be present.

As in the null context condition, we can derive more detailed predictions for individual constraint violations based on the set of constraints in (20). These predictions can be tested by further investigating the interactions listed above. Table 7 details which interactions will be used to test which constraints.

Firstly, we expect to find the context independent effects that were already discussed for the



Table 7: Interactions Used to Test the Constraint Set (Experiment 2, Context Condition)

Interaction	Constraints
ORD/DOU	DOUBLEALIGN
DOU/ACC	ACCENTFOCUS, DOUBLEGROUND
ACC/CON	ACCENTFOCUS
DOU/CON	DOUBLEGROUND
ORD/CON	GROUNDALIGN

null context condition: the constraint DOUBLEALIGN is violated in preverbal objects without doubling, i.e., we should find  $OclVS > OVS$ . Note, though, that there is the inherent conflict between DOUBLEGROUND and ACCENTFOCUS in stimuli with accented doubled objects which are therefore predicted to be less acceptable than doubled orders with subject accent, and than orders without doubling.

The constraints GROUNDALIGN, DOUBLEGROUND, ACCENTFOCUS, and ACCENTALIGN formalize the interaction of order, doubling, and accent with Information Structure. Under the optimality theoretic approach advocated in the present paper, we expect that the preference pattern for each context will reflect the optimal realization of the Information Structure specified by this context. The constraint GROUNDALIGN predicts that orders with non-peripheral ground constituents will be reduced in acceptability (see Experiment 1), while DOUBLEGROUND indicates that stimuli with doubled objects that are not part of ground should be dispreferred. ACCENTFOCUS predicts reduced acceptability for accented constituents that are not in focus, while ACCENTALIGN leads us to reduced acceptability for stimuli in a broad context that fail to carry the accent on the rightmost constituent.

The following predictions about the effects of GROUNDALIGN, DOUBLEGROUND, ACCENTFOCUS, and ACCENTALIGN can be made for each context. (Again, these predictions only deal with the effect of individual constraint violations; constraint interaction will be discussed separately in Section 4.5.)

*All Focus Context.* The predictions for the all focus context were already discussed in Section 4.2.1, based on the assumption that the null context and the all focus context behave in the same way. To recapitulate: no effects of GROUNDALIGN, DOUBLEGROUND, and ACCENTFOCUS are expected, as these constraints are vacuously satisfied in an all focus context.

ACCENTALIGN predicts that orders with the accent on the rightmost constituent are preferred, i.e.,  $svO > Svo$ ,  $ovS > Ovs$ , and  $vsO > vSo$ . For stimuli involving clitics, ACCENTALIGN predicts that  $Sclvo$ ,  $Oclvs$ , and  $clvSo$  will be reduced in acceptability.

*S Focus Context.* In the S focus context, the subject is in focus, while the object is part of ground. This means that GROUNDALIGN is satisfied by SVO, OVS, and VSO, and hence all three orders would be equally acceptable.

DOUBLEGROUND requires that doubled objects have to be interpreted as ground. This constraint is satisfied, as the S focus context marks the object as ground. Hence our constraint set predicts that doubled and non-doubled orders will be equally acceptable.

ACCENTFOCUS requires that accented constituents are interpreted as focus. This requirement is satisfied by orders with S accent, but violated by orders with O accent, because the S focus context

specifies the object as ground. Hence we predict that orders with S accent are more acceptable than orders with O accent.

Note that the scope of ACCENTALIGN is restricted to broad focus, hence no predictions can be derived for the S focus context.

*O Focus Context.* In the O focus context, the object is in focus, while the subject is part of ground. GROUNDALIGN is satisfied by SVO and OVS, but violated by VSO, where the ground constituent (the subject) is not peripheral. Hence we expect VSO to be reduced in acceptability compared to SVO and OVS.

DOUBLEGROUND requires that doubled objects have to be interpreted as ground. This constraint is violated by clitic doubled orders in the O focus context, as the object is focussed. Hence we predict clitic doubled orders to be less acceptable than doubled ones, which do not violate DOUBLEGROUND.

ACCENTFOCUS is met by orders with O accent, but violated by orders with S accent, as the O focus context specifies the subject as ground. Hence orders with O accent are expected to be more acceptable than S accented orders.

Again, no predictions could be derived from ACCENTALIGN, which only applies in broad focus contexts.

*V Focus Context.* In the V focus context, the verb is in focus, while the subject and the object are ground constituents. As discussed in Experiment 1, VSO incurs a violation of GROUNDALIGN, as the subject fails to be peripheral (i.e., appear either clause finally or clause initially). Hence we predict reduced acceptability for VSO compared to SVO and OVS.

No relevant prediction can be derived from ACCENTFOCUS, DOUBLEGROUND, and ACCENTALIGN. In the V focus context, all orders violate ACCENTFOCUS, as the accent is either on the subject or on the object (recall that V accent was not included in the stimulus set). DOUBLEGROUND, on the other hand, is satisfied by all orders, as the context marks the object as ground. ACCENTALIGN is not applicable, as we are dealing with a narrow focus context.

### 4.3. Method

#### 4.3.1. Subjects

Thirty-six native speakers of Greek participated in the experiment. The subjects were international students at the University of Edinburgh, Napier University, and Heriot-Watt University. The experiment was administered in the laboratory and subjects were paid for their participation. It was made sure that subjects were naive, i.e., they were neither linguists or students of linguistics. None of the subjects had previously participated in Experiment 1.

The data of three subjects were excluded as they were bilingual (by self-assessment). The data of one further subject were excluded as she was a speaker of Cypriot Greek. In addition, the data of two subjects were excluded after an inspection of the responses showed that they had not completed the task adequately. The data of one subject were lost due to a technical problem. This left twenty-nine subjects for analysis. Of these, 11 subjects were male, 18 female; six subjects were left-handed, 23 right-handed. The age of the subjects ranged from 20 to 37 years, the mean was 26.0 years.

All subjects were resident in Edinburgh at the time of the experiment. The overall time they had lived in an English-speaking environment ranged from 6 to 96 months, the mean was

29.1 months.

#### 4.3.2. *Materials*

*Training Materials.* The experiment included a set of training materials that were designed to familiarize subjects with the magnitude estimation task. The training set was larger than the one in Experiment 1, containing ten horizontal lines. The range of largest to smallest item was 1:10. The items were distributed evenly over this range, with the largest item covering the maximal window width of the web browser. A modulus item in the middle of the range was provided.

*Practice Materials.* A set of practice items was used to familiarize subjects with applying magnitude estimation to linguistic stimuli. The practice set consisted of ten sentences that were representative of the test materials: the practice items covered a wide spectrum of acceptability, ranging from fully acceptable to severely unacceptable. In addition, the set was balanced so that all word orders and accent patterns were equally represented. A modulus item in the middle of the range was provided.

*Test Materials.* For the experimental items, a full factorial design was used with word order (ORD), context (CON), clitic doubling (DOU), and accent placement (ACC) as the factors, yielding a total of  $ORD \times CON \times DOU \times ACC = 3 \times 5 \times 2 \times 2 = 60$  cells. Eight lexicalizations per cell were used, which resulted in a total of 480 stimuli.

A set of 48 fillers was used, designed to cover the whole acceptability range. Twelve items of each of the following four groups were used: no violation, case violation, phrase structure violation, and agreement violation. The set of fillers was balanced so that each word order and accent pattern used in the experimental items occurred equally often in the fillers. The context items for the fillers were also balanced to reflect the proportions in the experimental set.

Again, a modulus item in the middle of the range was provided (see Appendix B for a list of all experimental materials).

#### 4.3.3. *Recordings and Pretests*

*Recordings.* Practice and test materials were read by a male native speaker of Greek, who was unaware of the purpose of the experiment. The reader received brief training by the experimenters to make sure that he was able to produce the required accent patterns consistently. The experimental items were tape recorded and later sampled using the sound hardware of a Sparc Ultra 10 workstation. The sampling software used was Sun's Audiotool, with the sampling rate set at 8000 Hz. Questions and answers were recorded separately to exclude possible variations in the accent pattern caused by the context preceding a stimulus during recording.

*Intelligibility Pretest.* As the stimuli crucially relied on phonetically deficient elements (clitics), a pretest was carried out to insure that the stimuli were fully intelligible. Two native speakers of Greek were asked to judge the intelligibility of the stimuli. Under experimental conditions, they listened to the stimuli in random order. Each stimulus was presented once and the subject had to repeat it. The experimenter then compared the repetition to a written version of the stimulus. All stimuli that were repeated incorrectly by at least one of the subjects were re-recorded and re-tested. The intelligibility pretest included all experimental items (i.e., the full practice and test sets, including contexts and fillers).

*Accent Uniformity Pretest.* As the stimuli crucially relied on accent placement, a pretest was carried out to ensure that the accent patterns were uniformly realized in each experimental condition. Two phonetically trained speakers of Greek (one native and one near-native) were asked to judge whether the accent realized in each experimental condition was uniform across items. Under experimental conditions, the subjects listened to each item in each condition as often as they liked. They were told which accent was supposed to be realized in which condition (S or O accent) and had to judge whether one or more items in the condition had diverging accent patterns. These items were then re-recorded and re-tested. The accent uniformity pretest included only the test items (i.e., contexts, fillers, and practice items were not tested).

#### 4.3.4. Procedure

Again, magnitude estimation was used as the experimental paradigm. Each subject took part in an experimental session that lasted approximately 45 minutes and consisted of a training phase, a practice phase, and an experimental phase. The experiment was self-paced, though response times were recorded to allow the data to be screened for anomalies.

The experiment was conducted in a laboratory on PCs. Netscape 4.0 under Windows 95 was used to administer the experiment. The browser established an Internet connection to the experimental server, which controlled the experiment using WebExp 2.1 (Keller et al., 1998).

*Instructions, Demographic Questionnaire, and Training Phase.* The instructions were the same as in Experiment 1, except that they were adapted for spoken stimuli. The demographic questionnaire and the training phase were the same as in Experiment 1

*Practice Phase.* This phase allowed subjects to practice magnitude estimation of linguistic acceptability using spoken stimuli. Items were presented to subjects over headphones. For each item, the subject had to click on a Play button to start the presentation of this item. After the item finished playing, the subject had to provide a numerical judgment over the computer keyboard. After pressing Return, a new Play button for the next item was displayed. Each item had to be played exactly once, and there was no possibility to change responses once Return had been pressed. No time limit was set for the responses.

Subjects first judged the modulus item, and then all the items in the training set. Items were presented in random order, with a new randomization being generated for each subject.

*Experimental Phase.* Presentation and response procedure in the experimental phase was the same as in the practice phase. A between-subjects design was used to administer the experimental stimuli: subjects in Group A judged non-contextualized stimuli, while subjects in Group B judged contextualized stimuli.

For Group A, two test sets were used: each set contained four lexicalizations for each of the cells in the design  $ORD \times DOU \times ACC$ , i.e., a total of 48 items. Lexicalizations were assigned to test sets using a Latin square design. For Group B, eight test sets were used: each set contained one lexicalization for each of the cells in the design  $ORD \times CON \times DOU \times ACC$ , a total of 48 items. Again, a Latin square was used to assign lexicalizations to test sets.

As in the practice phase, subjects first judged the modulus item. Then they listened to 96 test items: 48 experimental items and 48 fillers. Items were presented in random order, with a new randomization being generated for each subject. Each experimental subject was randomly assigned

to a group and a lexicalization; 12 subjects were assigned to Group A, 17 to Group B. Instructions, examples, training items, and fillers were adapted for Group B to take context into account.

#### 4.4. Results

As in Experiment 1, the data were normalized by dividing each judgment by the modulus value and by log-transforming the result. All analyses were conducted on the normalized, log-transformed judgments.

Recall that all the figures in this paper display means of normalized, log-transformed judgments, together with standard errors. Appendix C contains the descriptive statistics for the experimental results.

##### 4.4.1. Null Context Condition

Separate ANOVAs were conducted for the context and the null context condition. The mean judgments for the null context condition are graphed in Figure 5. An ANOVA revealed a significant main effect of word order ( $F_1(2, 22) = 11.87, p < .0005$ ;  $F_2(2, 14) = 13.70, p = .001$ ). Significant main effects of clitic doubling ( $F_1(1, 11) = 13.87, p = .003$ ;  $F_2(1, 7) = 24.56, p = .002$ ) and accent placement were also present ( $F_1(1, 11) = 10.81, p = .007$ ;  $F_2(1, 7) = 19.20, p = .003$ ).

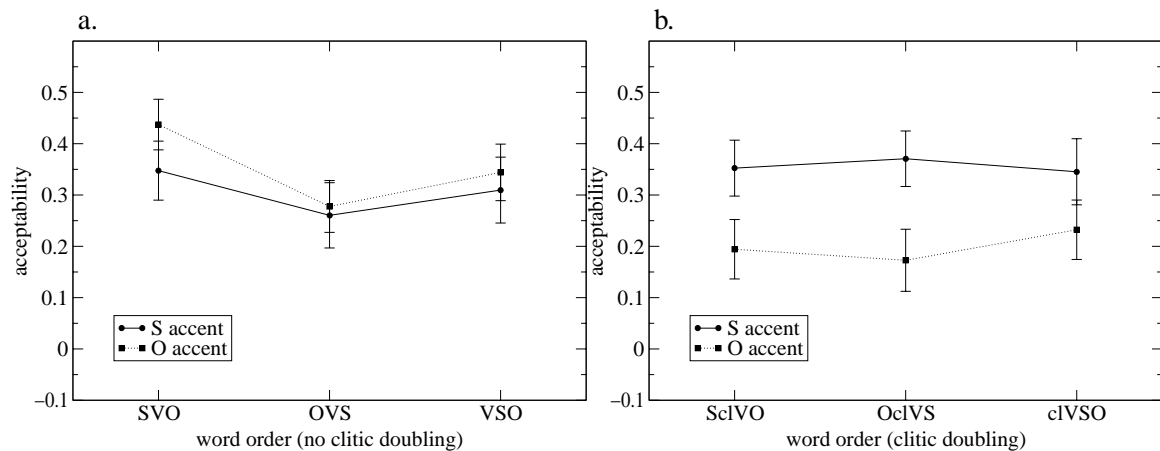


Figure 5. Word Order Preferences in the Null Context (Experiment 2)

As predicted, an interaction between word order and clitic doubling was found ( $F_1(2, 22) = 7.00, p = .004$ ;  $F_2(2, 14) = 15.77, p < .0005$ ), indicating that clitic doubling affects the acceptability of certain word orders. We also found an interaction between clitic doubling and accent ( $F_1(1, 11) = 27.70, p < .0005$ ;  $F_2(1, 7) = 46.72, p < .0005$ ). This interaction was predicted on the basis of the unacceptability of accented clitic doubled objects. Finally, there was an interaction of word order and accent ( $F_1(2, 22) = 5.33, p = .013$ ;  $F_2(2, 14) = 4.44, p = .032$ ). This is in line with the prediction that some word orders prefer S accent, while others prefer O accent. The three-way interaction of word order, clitic doubling, and accent placement failed to be significant.

Post-hoc Tukey tests were carried out on the interactions to test the predictions of individual constraints, in line with the schema in Table 6. The Tukey test for the ORD/DOU interaction allows us to assess the validity of DOUBLEALIGN, which predicts that OVS (violating DOUBLEALIGN)

should be less acceptable than SVO and VSO, while all clitic doubled orders should be equally acceptable. This prediction was borne out: OVS was significantly less acceptable than SVO ( $\alpha < .01$ ) and VSO (by items only,  $\alpha < .01$ ). The Tukey test also showed that SVO was more acceptable than VSO (by items only,  $\alpha < .01$ ), which was unexpected. On the other hand, the orders OclVS, SclVO, clVSO, were not significantly different from each other, in line with our predictions.

It is worth noting that the results in Figure 5 support our formulation of DOUBLEALIGN. The theoretical literature on Greek associates the requirement that preverbal objects should be doubled only with ground objects (Tsimpli, 1995; Tsipakou, 1998). No such restriction is assumed for focused preverbal objects. In contrast, our formulation of DOUBLEALIGN does not make any reference to the discourse function of the preverbal object. If this constraint were to apply only on ground preverbal objects, then Ovs should be much better than ovS, contrary to the results shown in Figure 5, where ovS and Ovs receive the same rating.

Furthermore, we predicted reduced acceptability for orders with object accent and clitic doubling, as these orders incur a conflict of DOUBLEGROUND and ACCENTFOCUS. This can be tested by performing a Tukey test on the DOU/ACC interaction. As predicted, we found that orders with O accent and doubling were significantly less acceptable than orders with S accent and doubling ( $\alpha < .01$ ), orders with S accent without doubling ( $\alpha < .01$ ), and orders with O accent without doubling ( $\alpha < .01$ ). As expected, there were no significant differences between non-doubled orders with S accent, non-doubled orders with O accent, and doubled orders with S accent.

Finally, we conducted a Tukey test on the ORD/ACC interaction to validate the constraint ACCENTALIGN, which requires that accented constituents have to be right peripheral. This predicts that svO should be preferred over Svo, ovS over Ovs, and vsO over vSo. The Tukey test showed that the preference ovS > Ovs was significant ( $\alpha < .01$ ), but failed to find a difference between svO and Svo, and between vsO and vSo.

#### 4.4.2. Context Condition

The mean judgments for the context condition are graphed in Figures 6–9. The ANOVA for the context condition yielded the same general picture as in the non-context condition: significant main effects of word order ( $F_1(2, 32) = 11.42, p < .0005; F_2(2, 14) = 8.27, p = .004$ ) and clitic doubling ( $F_1(1, 16) = 20.71, p < .0005; F_2(1, 7) = 17.01, p = .004$ ) were found. Accent, however, failed to reach significance. A main effect of context was also discovered ( $F_1(3, 48) = 11.55, p < .0005; F_2(3, 21) = 28.78, p < .0005$ ).

As in the null context condition, we found an interaction of word order and clitic doubling ( $F_1(2, 32) = 6.88, p = .003; F_2(2, 14) = 11.57, p = .001$ ), clitic doubling and accent ( $F_1(1, 16) = 23.44, p < .0005; F_2(1, 7) = 24.13, p = .002$ ), and word order and accent ( $F_1(2, 32) = 6.28, p = .005; F_2(2, 14) = 5.20, p = .020$ ).

The ANOVA also demonstrated an interaction of accent and context ( $F_1(3, 48) = 26.36, p < .0005; F_2(3, 21) = 33.10, p < .0005$ ), showing that accent placement has an information structural effect, as predicted. We also discovered an interaction of clitic doubling and context ( $F_1(3, 48) = 15.16, p < .0005; F_2(3, 21) = 10.87, p < .0005$ ), which confirms that clitic doubling interacts with Information Structure. In addition, we found a significant interaction of word order and context ( $F_1(6, 96) = 7.72, p < .0005; F_2(6, 42) = 7.12, p < .0005$ ). This confirms the finding in Experiment 1 that word order preferences are subject to context effects. All other interactions failed to reach significance.

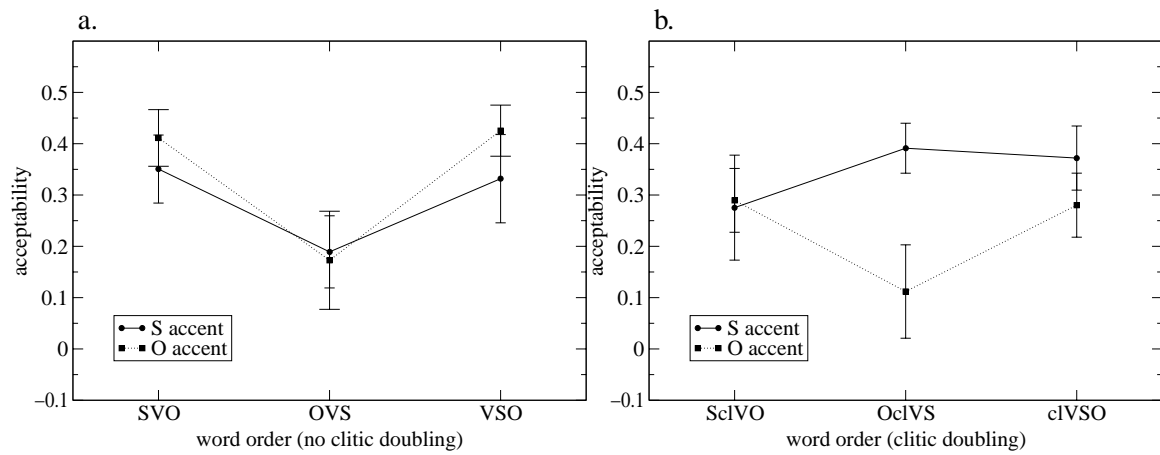


Figure 6. Word Order Preferences in the All Focus Context (Experiment 2)

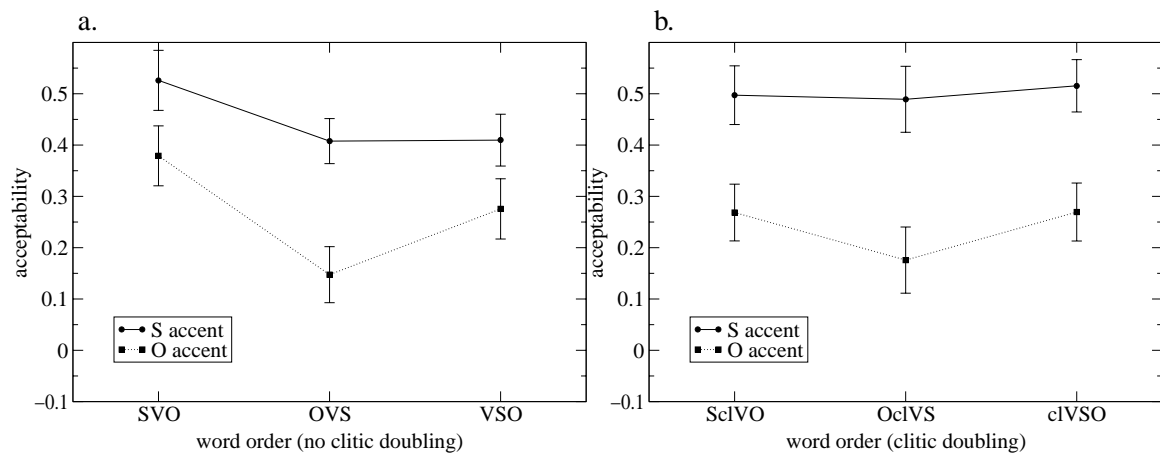


Figure 7. Word Order Preferences in the S Focus Context (Experiment 2)

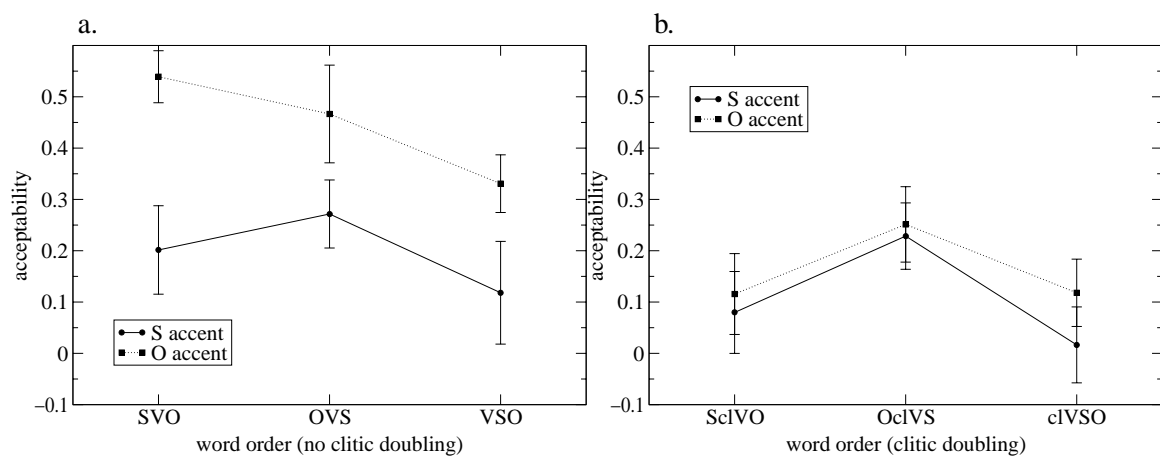


Figure 8. Word Order Preferences in the O Focus Context (Experiment 2)

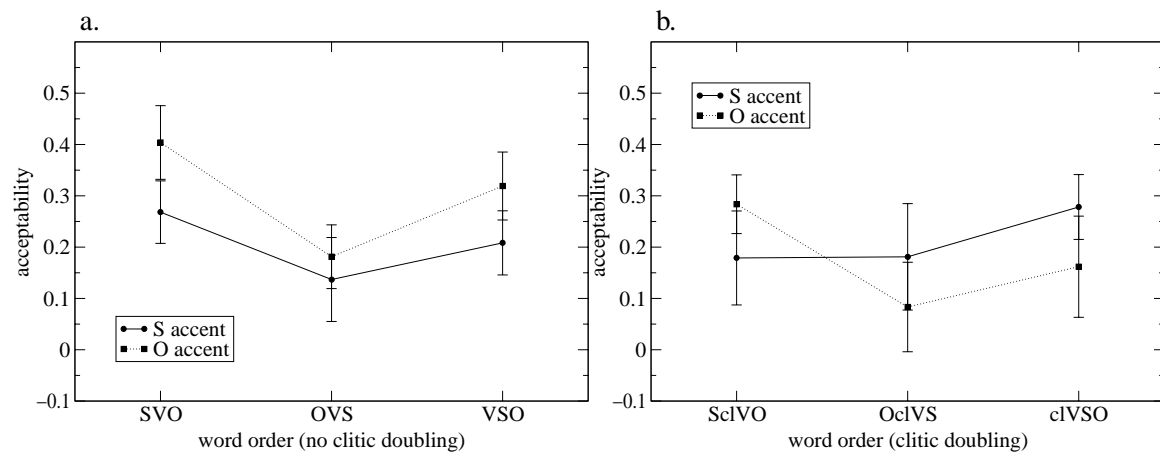


Figure 9. Word Order Preferences in the V Focus Context (Experiment 2)

Post-hoc Tukey tests were carried out on the interactions to test the predictions of individual constraints (see Table 7 for details). We will first report the results for context independent constraints. A Tukey test on the ORD/DOU interaction was conducted to test the constraint DOUBLEALIGN. As predicted, OVS was less acceptable than SVO ( $\alpha < .01$ ) and VSO (by items only,  $\alpha < .05$ ). We also found the unexpected preference SVO > VSO (by items only,  $\alpha < .01$ ). The three clitic doubled orders ScIVO, OcIVS, and cIVSO did not differ significantly in acceptability, which is in line with our predictions.

The second context independent prediction was that orders with object accent and clitic doubling should be less acceptable than other orders, as they incur a conflict of DOUBLEGROUND and ACCENTFOCUS. As in the null context condition, we performed a Tukey test on the DOU/ACC interaction to test this prediction. We found that orders with O accent and doubling were significantly less acceptable than orders with S accent and doubling ( $\alpha < .01$ ), orders with S accent without doubling ( $\alpha < .01$ ), and orders with O accent without doubling ( $\alpha < .01$ ). As expected, there were no significant differences between non-doubled orders with S accent, non-doubled orders with O accent, and doubled orders with S accent.

The constraints GROUNDALIGN, DOUBLEGROUND, ACCENTFOCUS, and ACCENTALIGN make specific predictions for each context, which we discuss separately below.

*All Focus Context.* In this context, GROUNDALIGN is vacuously satisfied. Therefore we predicted that there should be no difference between the orders SVO, OVS, and VSO. To verify this prediction, we conducted a post-hoc test on the interaction ORD/CON. There was no significant difference between SVO and VSO, but we found that OVS was significantly less acceptable than both SVO (by items only,  $\alpha < .05$ ) and VSO ( $\alpha < .01$ ), contrary to what was expected. Figure 6a provides an explanation for this finding: OVS without doubling violates the constraint DOUBLEALIGN, which greatly reduces its acceptability. This effect is not present in clitic doubled stimuli (see Figure 6b).

As mentioned earlier (Section 1.2), an all focus context can accommodate a wider range of Information Structures. In particular, doubled objects, characteristically associated with a ground interpretation, are felicitous in an all focus context (see examples 11 and 12 and the relevant dis-



cussion in Section 1.2). Hence DOUBLEGROUND was expected to induce no effects in the all focus context. The Tukey test on the interaction DOU/CON confirmed this by failing to indicate a significant difference between doubled and non-doubled orders.

ACCENTALIGN predicted that orders with the accent on the rightmost constituent are preferred. We used planned comparisons to test this prediction (post-hoc tests could not be performed as there was no three-way interaction ACC/CON/ORD). Adjusting the significance level using the Bonferroni method, we set  $p = .017$ , as three comparisons were carried out.

According to ACCENTALIGN, svO should be preferred over Svo, ovS over Ovs, and vsO over vSo. A set of one-way ANOVAs showed that the preference  $ovS > Ovs$  was significant (by items only,  $F_1(1, 17) = 4.74$ ,  $p = .045$ ;  $F_2(1, 7) = 20.17$ ,  $p = .003$ ), but failed to find a difference between svO and Svo, and between vsO and vSo. These results mirrors the ones obtained in the null context, and constitute a partial confirmation of ACCENTALIGN.

*S Focus Context.* In the S focus context, all three orders, SVO, OVS, and VSO, satisfy the constraint GROUNDALIGN and are, therefore, expected to show no significant differences in acceptability. This prediction was borne out by a Tukey test on the interaction ORD/CON (see Figure 7).

DOUBLEGROUND requires that doubled objects have to be ground. This requirement is satisfied in an S focus context, where objects are marked as ground elements. Hence doubled and non-doubled orders should be equally acceptable. In line with this prediction, the Tukey test on the interaction DOU/CON failed to find a significant difference between doubled and non-doubled orders.

The constraint ACCENTFOCUS requires that accented constituents have to be in focus. For the S focus context, this predicts that orders with S accent should be more acceptable than orders with O accent. A Tukey test on the ACC/CON interaction confirmed this expectation ( $\alpha < .01$ ) (see also Figure 7).

Recall that the scope of ACCENTALIGN is restricted to broad focus contexts, hence no predictions could be derived for the S focus context. (The same holds for the other narrow focus contexts, i.e., O focus and V focus.)

*O Focus Context.* In this context, GROUNDALIGN is satisfied by SVO and OVS, but violated by VSO. Hence VSO should be reduced in acceptability compared to the verb medial orders (see Figure 8). A Tukey test on the ORD/CON interaction confirmed that VSO was less acceptable than OVS ( $\alpha < .01$ ). The  $SVO > VSO$  preference, however, failed to reach significance.

In O focus, orders with clitic doubling violate DOUBLEGROUND and hence are predicted to be less acceptable than non-doubled orders. The Tukey test on the interaction DOU/CON confirmed this prediction ( $\alpha < .01$ ).

In the O focus context, the constraint ACCENTFOCUS predicts that orders with O accent should be more acceptable than orders with S accent. This prediction was borne out by the Tukey test on the ACC/CON interaction ( $\alpha < .01$ ) (see also Figure 8).

*V Focus Context.* In this context, GROUNDALIGN predicts reduced acceptability for VSO compared to SVO and OVS. This prediction could not be confirmed by the Tukey test on the ORD/CON interaction, which failed to find a difference between VSO and SVO, and between VSO and OVS. However, we found the significant preference  $SVO > OVS$  ( $\alpha < .01$ ). This is probably due to the fact that OVS (without doubling) violates DOUBLEALIGN (see also Figure 9).

DOUBLEGROUND predicts that doubled and non-doubled orders are equally acceptable as the object is part of ground in the V focus context. In line with this, the Tukey test on the DOU/CON interaction failed to find a difference between doubled and non-doubled orders.

Note that in the V focus context, ACCENTFOCUS is always violated (as V accent was not included in our stimulus set). This explains why all orders receive fairly low acceptability scores compared to the optimal orders in the O focus and S focus contexts. Furthermore, it seems that the overall acceptability pattern is fairly similar to the one obtained in the all focus context (compare Figures 6 and 9).

#### 4.5. Optimality Theoretic Modeling

Experiment 2 confirmed the results of Experiment 1 by providing additional evidence for the validity of the constraints DOUBLEALIGN and GROUNDALIGN. DOUBLEALIGN, which requires preverbal objects to be doubled, is satisfied by SVO and VSO, but violated by OVS. This was in line with the experimental data: for non-doubled orders, we found that SVO and VSO were significantly more acceptable than OVS. The doubled orders OcIVS, ScIVO, and cIVSO did not differ in acceptability, in line with the predictions of DOUBLEALIGN. The constraint GROUNDALIGN requires ground constituents to be sentence peripheral. This was consistent with the findings for the S focus context, where SVO, OVS, and VSO were not significantly different. Moreover, in the O context we found that VSO was less acceptable than OVS, as predicted by GROUNDALIGN.

Experiment 2 also tested three additional constraints: DOUBLEGROUND, ACCENTFOCUS, and ACCENTALIGN. DOUBLEGROUND states that doubled objects have to be interpreted as ground. This constraint is satisfied in the all focus, S focus, and V focus context. Doubled and non-doubled stimuli were equally acceptable in these contexts, as predicted. In O focus, doubled stimuli violate DOUBLEGROUND and were less acceptable than non-doubled ones. The constraint ACCENTFOCUS requires that accented constituents have to be interpreted as focus; this was confirmed in the S focus context, where stimuli with S accent were more acceptable than stimuli with O accent. In the O focus context, the pattern was reversed. ACCENTALIGN predicts that accented constituents have to be right peripheral in broad focus. Tendencies in line with the predictions of ACCENTALIGN could be observed in the null context and in the all focus context, but the overall evidence for ACCENTALIGN was rather weak.

As in Experiment 1, we will use the experimental data not only to validate our constraint set, but also to establish a ranking between the constraints. This can be achieved by using suboptimal ranking arguments, as defined in (19). To rank five constraints, four ranking arguments are required, which are detailed below. As in Experiment 1, we use one-way ANOVAs (planned comparisons) to determine if the acceptability difference for a given ranking argument is significant. A significance threshold of  $p = .0125$  will be assumed (Bonferroni adjustment for four comparisons).

Let us first consider the ranking of ACCENTALIGN with respect to DOUBLEALIGN. In the all focus context (see Figure 6a), we can compare Svo (which violates ACCENTALIGN, but satisfies DOUBLEALIGN) with ovS (which violates DOUBLEALIGN, but satisfies ACCENTALIGN). The preference pattern is Svo > ovS. Another minimal pair is afforded by ovS and vSo, where we observe the ranking vSo > ovS. Both minimal pairs show that a violation of DOUBLEALIGN is more serious than a violation of ACCENTALIGN. This is confirmed by a one-way ANOVA on the combined data from both minimal pairs (by subjects only,  $F_1(1, 16) = 9.26$ ,  $p = .008$ ;  $F_2(1, 7) = 6.78$ ,  $p = .035$ ). We, thus, conclude that DOUBLEALIGN  $\gg$  ACCENTALIGN is the correct ranking.

The next ranking argument involves the constraints GROUNDALIGN and DOUBLEALIGN.

In the O focus context (see Figure 8a) Ovs and vsO form a minimal pair, as Ovs violates DOUBLEALIGN, but satisfies GROUNDALIGN, whereas vsO violates GROUNDALIGN, but satisfies DOUBLEALIGN. The acceptability pattern is Ovs > vsO. Another minimal pair is provided by ovS and vSo (also in the O context), where ovS violates DOUBLEALIGN, while vSo violates GROUNDALIGN, and the preference is ovS > vSo. A one-way ANOVA on the combined data from the two minimal pairs shows that a GROUNDALIGN violation is more serious than a DOUBLEALIGN violation (by subjects only,  $F_1(1, 16) = 8.65$ ,  $p = .010$ ;  $F_2(1, 7) = 7.80$ ,  $p = .027$ ). This confirms the ranking GROUNDALIGN  $\gg$  DOUBLEALIGN that was already established in Experiment 1.

Regarding the ranking of DOUBLEGROUND and GROUNDALIGN, the O focus context (see Figure 8) affords a minimal pair: vsO violates GROUNDALIGN, while sclvO violates DOUBLEGROUND. The orders vsO and Oclvs constitute a further minimal pair. We observe the preferences vsO > sclvO and vsO > Oclvs. An ANOVA on the combined data confirms that a violation of DOUBLEGROUND is more serious than a violation of GROUNDALIGN (by subjects only,  $F_1(1, 16) = 23.56$ ,  $p < .0005$ ;  $F_2(1, 7) = 6.28$ ,  $p = .041$ ), which demonstrates that DOUBLEGROUND  $\gg$  GROUNDALIGN is the correct ranking.

Now consider the ranking of ACCENTFOCUS with respect to DOUBLEGROUND. Minimal pairs are again provided by the O focus context (see Figure 8), where Svo violates ACCENTFOCUS, while sclvO violates DOUBLEGROUND. The orders vSo and clvsO constitute another minimal pair. An ANOVA on the combined data from the two minimal pairs fails to find a significant acceptability difference between the orders that violate ACCENTFOCUS and the orders that violate DOUBLEGROUND, which leads to the conclusion that both constraints are equally ranked.

To summarize, we have established the following constraint hierarchy based minimal pairs that were present in the data from Experiment 2:

$$(27) \{ \text{DOUBLEGROUND, ACCENTFOCUS} \} \gg \text{GROUNDALIGN} \gg \text{DOUBLEALIGN} \gg \text{ACCENTALIGN}$$

The hierarchy in (27) can now be used to compute a constraint violation profile for each context, based on the suboptimality hypothesis (see (18)). It yields the constraint profiles in Tables 8–11 for each of the context investigated in the present experiment.

Each of these constraint profiles predicts a grammaticality order on a given set of candidate structures. This order can then be compared against the acceptability order that was obtained experimentally for the same set of structures. As in Experiment 1, we determined the acceptability order by converting the average judgments to ranks, ignoring differences that are smaller than one standard error. Using correlation analysis, we can then determine how well the theoretically predicted order matches the experimentally found one (see Section 3.5 for details).

The profile in Table 8 predicts the grammaticality order in (28a) for both the null and the all focus context. Recall that we assume that doubled objects are felicitous in the null context and the all focus context (see Section 1.2), hence these contexts do not incur violations of DOUBLEGROUND.<sup>11</sup> Experimentally, we found the order in (28b) for the null context, and the order in (28c) in the all focus context.

<sup>11</sup>Nothing in our model distinguishes narrow focus contexts from broad focus contexts, which makes our decision to disregard the effect of DOUBLEGROUND in all focus seem ad hoc (from a model-internal point of view). A theoretically motivated account of the differences between broad and narrow focus contexts based on pragmatic principles is possible, but falls outside the scope of the present paper.

Table 8: Constraint Profile for the Null Context and the All Focus Context (Experiment 2)

all focus	ACCF	DOUG	GAGN	DOUAGN	ACCAGN
Svo					*
ovS				*	
vSo					*
svO					
Ovs				*	*
vsO					
ScIvo					*
oclvs					
clvSo					*
sclvO		*			
Oclvs		*			*
clvsO		*			

(28) **Null Context and All Focus Context**

- a. **Prediction:** {svO, vsO, oclvS} > {Svo, vSo, ScIvo, clvSo} > ovS > Ovs > {sclvO, clvsO} > Oclvs
- b. **Result Null Context:** svO > {vsO, oclvS, Svo, ScIvo, clvSo} > {vSo, ovS, Ovs} > clvsO > {sclvO, Oclvs}
- c. **Result All Focus Context:** {svO, vsO, oclvS, clvSo} > {Svo, vSo} > {ScIvo, sclvO, clvsO} > {ovS, Ovs} > Oclvs

For the null context, a high correlation between the predicted order (see (28a)) and the actual acceptability order (see (28b)) was obtained ( $\rho(12) = .91, p < .0005$ ). The main mismatch was that vsO and vSo were less acceptable than predicted. (Note that we found a reduced acceptability of VSO also in Experiment 1.) Also, the acceptability of oclvS was slightly lower than expected.

The correlation for the all focus context (see (28c)) was also substantial ( $\rho(12) = .86, p < .0005$ ). Surprisingly, two of the orders with accented doubled objects were more acceptable than expected; sclvO and clvsO were predicted to be clearly dispreferred over ScIvo and clvSo, as they violate the highly ranked constraint DOUBLEGROUND. Furthermore, clvSo was more acceptable than predicted, while ScIvo was less acceptable than predicted.

We also found a correlation between the acceptability order for the all focus context and the null context ( $\rho(12) = .76, p = .002$ ). This confirms our hypothesis that even when faced with isolated sentences, speakers make implicit assumptions about Information Structure—they assume an all focus context.

We will now discuss the model fit for the narrow focus contexts. For the S focus context, Table 9 predicts the grammaticality order in (29a). The order obtained experimentally is listed in (29b).

(29) **S Focus Context**

- a. **Prediction:** {Svo, vSo, ScIvo, oclvS, clvSo} > ovS > {svO, vsO, sclvO, Oclvs, clvsO} > Ovs

Table 9: Constraint Profile for the S Focus Context (Experiment 2)

S focus	ACCF	DOUG	GAGN	DOUAGN	ACCAGN
Svo					
ovS				*	
vSo					
svO	*				
Ovs	*			*	
vsO	*				
Scuvo					
oclvS					
clvSo					
sclvO	*				
Oclvs	*				
clvsO	*				

- b. **Result:** {Svo, Scuvo, oclvS, clvSo} > {vSo, ovS} > svO > {vsO, sclvO, clvsO} > {Oclvs, Ovs}

Again, there was a high correlation between the predicted order in (29a) and the order experimental order in (29b) ( $\rho(12) = .93, p < .0005$ ). Most orders were ranked correctly, with the exception of vSo, which was less acceptable than expected. Note that a similar effect was observed also in Experiment 1, where VSO was not as acceptable as predicted by its violation profile. Two other mismatches occurred: the order svO was slightly more acceptable than predicted, while Oclvs was less acceptable than expected.

Table 10: Constraint Profile for the O Focus Context (Experiment 2)

O focus	ACCF	DOUG	GAGN	DOUAGN	ACCAGN
Svo	*				
ovS	*			*	
vSo	*		*		
svO					
Ovs				*	
vsO			*		
Scuvo	*	*			
oclvS	*	*			
clvSo	*	*	*		
sclvO		*			
Oclvs		*			
clvsO		*	*		

Now consider Table 10 which states the constraint profile for the O focus context. The corre-

sponding grammaticality order is given in (30a), while the acceptability order obtained experimentally is listed in (30b).

(30) **O Focus Context**

- a. **Prediction:** svO > Ovs > vsO > {Svo, sclvO, Oclvs} > ovS > {vSo, clvsO} > {Scuvo, oclvS} > clvSo
- b. **Result:** svO > Ovs > vsO > {Oclvs, oclvS, ovS} > Svo > {sclvO, vSo, clvsO} > Scuvo > clvSo

A comparison between (30a) and (30b) yields a correlation that is smaller than in the S focus context, but still substantial ( $\rho(12) = .80, p = .001$ ). There are a number of mismatches between the two ranks. Most strikingly, the order oclvS was more acceptable than expected. Other minor mismatches are that the acceptability of ovS and Scuvo was slightly higher than expected, while the acceptability of sclvO was lower than expected. Also note that the six orders at the top of the acceptability range were predicted correctly (svO, Ovs, vsO, Svo, Oclvs), while the mismatches typically occur for seriously unacceptable orders.

Table 11: Constraint Profile for the V Focus Context (Experiment 2)

V focus	ACCF	DOUG	GAGN	DOUAGN	ACCAGN
Svo	*				
ovS	*			*	
vSo	*		*		
svO	*				
Ovs	*			*	
vsO	*		*		
Scuvo	*				
oclvS	*				
clvSo	*		*		
sclvO	*				
Oclvs	*				
clvsO	*		*		

Finally, consider Table 11, the constraint profile for the V focus context. The match for this context is not expected to be very good, as ACCENTFOCUS is violated by all the orders that were included in Experiment 2. Hence all the stimuli were of low acceptability. (31a) gives the grammaticality order predicted by our account for the V focus contest, while (31b) lists the acceptability order found experimentally.

(31) **V Focus Context**

- a. **Prediction:** {Svo, svO, Scuvo, oclvS, sclvO, Oclvs} > {ovS, Ovs} > {vSo, vsO, clvSo, clvsO}
- b. **Result:** svO > vsO > {Svo, clvSo, sclvO} > {ovS, vSo, Ovs, Scuvo, oclvS, clvsO} > Oclvs

We failed to find a correlation between (31a) and (31b) ( $\rho(12) = -.01, p = .485$ ). A manual comparison of the two order shows that the OT-model predicts some of the patterns in the data, such

as the reduced acceptability for *ovS* and *Ovs*, due to the violation of *DOUBLEALIGN*. Another example is the reduced acceptability of *vSo* and *vsO* compared to *Svo* and *svO*, which was due to the *GROUNDALIGN* violation incurred by *vSo* and *vsO*. Other than that, the model achieved only a poor fit with the experimentally obtained pattern.

To summarize, we were able to show that our optimality theoretic model accounts for the word order preferences elicited from native speakers. We demonstrated that the constraints in (20) together with the constraint hierarchy in (27) predict a set of grammaticality orders that match the acceptability orders obtained experimentally. Correlation analysis was used to test how well our model fits the data, and high correlation coefficients were obtained for all but the *V* focus context.

#### 4.6. Discussion

The results of the present experiment provide support for three aspects of the model put forward in this paper. Firstly, we were able to confirm the general hypothesis that word order, clitic doubling, and accent placement play an information structural role in a free word order language like Greek: we found significant interactions of word order and context, clitic doubling and context, accent and context.

Secondly, the experiment extended the results of Experiment 1 by providing evidence for a total of five grammatical constraints: the word order constraints *GROUNDALIGN*, the clitic doubling constraints *DOUBLEALIGN* and *DOUBLEGROUND*, and the accent constraints *ACCENTALIGN* and *ACCENTFOCUS* (see (20) for details). All of these constraints were well supported by the experimental findings, with the exception of *ACCENTALIGN*, which only manifested itself in weak tendencies. Further experimental data will be necessary to back up *ACCENTALIGN*. As this constraint only applies in broad focus contexts, experiments using *VP* focus stimuli are appropriate (these were not included in the present experiment).

Thirdly, the experiment extended the constraint hierarchy developed in Experiment 1 using suboptimal ranking arguments. The extended constraint hierarchy formalizes the interaction of syntactic and phonological constraints on Information Structure. Using a correlation analysis, we compared the grammaticality orders predicted by our model with the preference orders obtained experimentally. High correlations were achieved for all contexts, with the exception of the *V* focus context.

As mentioned in Section 1.2, word order is highly ambiguous in information structural terms. On the other hand, accent and doubling are unambiguously associated with focus and ground, respectively. We therefore predicted that violations of constraints on accent placement and doubling (for a given context) induce stronger effects than violations of word order preferences, a prediction which was confirmed by the experimental results. This fact is readily captured by the constraint rankings that we derived, as shown in the constraint hierarchy in (27), where the constraints *DOUBLEGROUND* and *ACCENTFOCUS* outrank the word order constraint *GROUNDALIGN*.

Another important finding of Experiment 2 is that an all focus context behaves like a null context (compare Figure 5 and Figure 6); a high correlation was obtained between the two preference patterns. This replicates the results of Experiment 1 for a wider range of context sensitive phenomena and for spoken stimuli, thus providing further support for the hypothesis that subjects make minimal contextual assumptions when they are exposed to isolated sentences: a null context is treated like an all focus context, which is what is expected under an information structural approach.

## 5. General Discussion

The work reported in this paper differs from standard theoretical approaches to Information Structure in both methodology and aims. Despite the differences though, it is worth discussing briefly how our results compare with the main hypotheses in the theoretical literature.

The literature on English has long acknowledged the major role of phonology in the realization of Information Packaging. English relies on accent placement and only rarely on syntax (e.g., topicalization, left dislocation) for discourse purposes. On the other hand, the literature on free word order languages like Catalan, Finnish, Greek, or Hungarian has emphasized the role of word order and has understated the role of accent placement. Our results from Greek, a standard example of a language exploiting word order for discourse purposes, indicate that, at least in this language, word order plays only a secondary role, whereas accent placement, as in English, is the most significant cue for signaling the ground-focus partition.

A detailed evaluation of the relevant theoretical analyses is beyond the scope of the present study. However, some observations can be made by comparing our findings to existing theoretical accounts of Information Structure in free word order languages. According to the dominant approach, languages like Greek or Hungarian encode Information Structure using specific phrase structure configurations, viz., the Focus Phrase (FP) and the Topic Phrase (TP). These phrases host focused and topic constituents, respectively (Kiss, 1995b; Tsimpli, 1995). Such languages are thus described as *discourse configurational*, in contrast to English, which has no discourse configurations. Analyses in this framework are typically concerned with the syntactic properties of constructions yielding the different word orders (focus movement, topicalization, clitic left dislocation) and stipulate distinct syntactic operations for each of them while paying little attention to accent placement. These approaches seem to imply that the syntactic operations yielding the different word orders play a more important role than accent placement, a claim that is incompatible with our results from Experiment 2. We found that, at least in Greek, word order (focus movement, topicalization, clitic left dislocation) plays only a secondary role in marking Information Structure; word order preferences can be overridden by phonological constraints.

On the other hand, our results lend indirect support to alternative analyses that do not associate the various orders with discourse configurations. Instead, a small number of syntactic constructions, underspecified for a particular discourse function, interact with accent placement in order to define the Information Structure of a sentence (Alexopoulou, 1998; Vallduví, 1992; Vallduví & Engdahl, 1996). An important aspect distinguishing such analyses from discourse configurational ones is that they assume that languages such as English and Greek share essentially the same set of information structural constraints. This is confirmed by our experimental findings, which suggest that Greek and English are much closer, at least with respect to the role of phonology, than suggested by discourse configurational approaches.

Partly, the explanation for the weak information structural role of word order lies in its ambiguity; for example the order SVO does not violate any constraint in any context, i.e., it is compatible with all Information Structures investigated in our experiments. However, even when word order is not ambiguous and does violate ordering constraints, the resulting preferences are weak. For example, VSO is not felicitous in an O focus context (because the ground NP is not dislocated to a peripheral position). However, the violation of the ordering constraint is less serious than a violation of accent placement (compare vsO and ovS in Figure 8a). On the other hand, there are cases where syntactic constraints on Information Structure are as strong as phonological constraints: violating



the requirement that focused NPs cannot be doubled is as serious as violating the requirement that accent falls on the focused constituent (see Figure 8 for the interaction between doubling and accent placement in O focus).

It seems that the ‘strength’ of a constraint depends on the extent to which it reflects an (unambiguous) grammaticalization of an information structural function. For example, accent and doubling could be viewed as grammaticalizations of focus and ground, respectively, whereas word order does not reflect such a grammaticalization. Note also that word order can be affected by independent factors such as the heaviness of the dislocated constituent and other discourse properties such as definiteness.

Issues of constraint strength are addressed neither by the discourse configurational nor by the alternative non-configurational approaches of Alexopoulou (1998), Vallduví (1992), and Vallduví and Engdahl (1996). This is partly due to the fact that these analyses rely on informal, introspective data which do not readily allow claims about constraint strength. Another reason is that none of the existing approaches (be it configurational or non-configurational) makes use of the concept of constraint competition that constitutes one of the central new developments in linguistic frameworks such as Optimality Theory. The present paper attempted to demonstrate how both problems can be overcome. We used experimental acceptability data to obtain information about the relative strength of violations of constraints on Information Structure. By making use of an extended version of OT we were able to formalize the intuition that constraint competition is responsible for the gradient acceptability patterns that emerge in the interaction of word order, accent placement and clitic doubling.

While these findings provide initial evidence for the hypothesis that interesting linguistic insights can be gained from investigating suboptimal structures, this result is limited to the effects of certain information structural constraints in Greek. In particular, our account lacks a crosslinguistic dimension. The OT approach to crosslinguistic variation as constraint re-ranking entails that the constraint hierarchy in (27) is language specific, i.e., that other languages will rank the same constraint differently. A language like Hungarian, where some contexts impose stricter restrictions on word order (for example, narrow foci are obligatorily preverbal, unlike in Greek) would be an interesting test case for this prediction.

Another important result of this paper is that the all focus and the null context yield the same general acceptability pattern. This indicates that so-called default constraints (that hold in the absence of context) are in fact the constraints that hold in an all focus context and can be explained by Information Packaging. Thus the behavior of isolated and contextualized sentences can receive a unified explanation under an information structural approach.

A main innovation of the present paper is that it models gradient grammaticality based on an extended version of Optimality Theory (instead of just accounting for binary grammaticality, as standard OT does). Such an approach raises a number of new questions, including the problem of parameter estimation. How can we determine the parameters of the model (i.e., the constraint ranks) such that an optimal fit with the experimental data is achieved? Keller (2000) deals with this problem and presents an algorithm that computes the constraint hierarchy that best accounts for a given set of acceptability judgments. This approach is based on a variant of OT that assumes that constraint violations are additive (the reader is referred to Keller, 2000, for details).

## References

- Agouraki, Y. (1993). *Spec-head licensing: The scope of the theory*. Unpublished doctoral dissertation, University College London.
- Alexopoulou, T. (1998). *The syntax of discourse functions in Greek: A non-configurational approach*. Unpublished doctoral dissertation, University of Edinburgh.
- Anagnostopoulou, E. (1994). *Clitic dependencies in Modern Greek*. Unpublished doctoral dissertation, University of Salzburg.
- Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, 72(1), 32–68.
- Birch, S., & Clifton, C., Jr. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, 38(4), 365–391.
- Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.
- Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, 32(1).
- Bolinger, D. L. (1978). Asking more than one thing at a time. In H. Hiz (Ed.), *Questions* (pp. 87–106). Dordrecht: Reidel.
- Bolinger, D. L. (1989). *Intonation and its uses: melody in grammar and discourse*. Stanford: Stanford University Press.
- Bresnan, J. (2000). Optimal syntax. In J. Dekkers, F. van der Leeuw, & J. van de Weijer (Eds.), *Optimality Theory: Phonology, syntax, and acquisition* (pp. 334–385). Oxford: Oxford University Press.
- Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–55). New York: Academic Press.
- Chafe, W. L. (1983). *Meaning and the structure of language*. Chicago: University of Chicago Press.
- Choi, H.-W. (1996). *Optimizing structure in context: Scrambling and information structure*. Unpublished doctoral dissertation, Stanford University.
- Corley, M., Keller, F., & Scheepers, C. (2000). *Conducting psycholinguistic experiments over the world wide web*. (in preparation)
- Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks, CA: Sage Publications.
- Dekkers, J., Leeuw, F. van der, & Weijer, J. van de (Eds.). (2000). *Optimality Theory: Phonology, syntax, and acquisition*. Oxford: Oxford University Press.
- Grimshaw, J. (1997). Projection, heads, and optimality. *Linguistic Inquiry*, 28, 373–422.
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English, part II. *Journal of Linguistics*, 3, 199–244.
- Hayes, B. P. (1997). *Four rules of inference for ranking argumentation*. (Unpubl. ms., Department of Linguistics, University of California, Los Angeles)
- Hayes, B. P. (2000). Gradient well-formedness in Optimality Theory. In J. Dekkers, F. van der Leeuw, & J. van de Weijer (Eds.), *Optimality Theory: Phonology, syntax, and acquisition* (pp. 88–120). Oxford: Oxford University Press.
- Horvath, J. (1995). Structural focus, structural case, and the notion of feature-assignment. In K. É. Kiss (Ed.), *Discourse configurational languages* (pp. 28–64). Oxford: Oxford University Press.

- Johnson-Laird, P. N., & Savary, F. (1999). Illusory inference: A novel class of erroneous deductions. *Cognition*, 71(3), 191–229.
- Keller, F. (1997). Extraction, gradedness, and optimality. In A. Dimitriadis, L. Siegel, C. Surek-Clark, & A. Williams (Eds.), *Proceedings of the 21st Annual Penn Linguistics Colloquium* (pp. 169–186). Department of Linguistics, University of Pennsylvania.
- Keller, F. (2000). *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. Unpublished doctoral dissertation, University of Edinburgh.
- Keller, F., Corley, M., Corley, S., Konieczny, L., & Todirascu, A. (1998). *WebExp: A Java toolbox for web-based psychological experiments* (Technical Report No. HCRC/TR-99). University of Edinburgh: Human Communication Research Centre.
- Kiss, K. É. (Ed.). (1995a). *Discourse configurational languages*. Oxford: Oxford University Press.
- Kiss, K. É. (1995b). NP movement, operator movement and scrambling in Hungarian. In K. É. Kiss (Ed.), *Discourse configurational languages* (pp. 207–243). Oxford: Oxford University Press.
- Ladd, R. D. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- Legendre, G., Wilson, C., Smolensky, P., Homer, K., & Raymond, W. (1995). Optimality and Wh-Extraction. In J. Beckman, L. W. Dickey, & S. Urbanczyk (Eds.), *Papers in Optimality Theory* (pp. 607–636). University of Massachusetts, Amherst.
- Lodge, M. (1981). *Magnitude scaling: Quantitative measurement of opinions*. Beverly Hills, CA: Sage Publications.
- Mehler, J. (1999). Editorial. *Cognition*, 71(3), 187–189.
- Müller, G. (1999). Optimality, markedness, and word order in German. *Linguistics*, 37(5), 777–818.
- Philippaki-Warbuton, I. (1985). Word order in Modern Greek. *Transactions of the Philological Society*, 113–143.
- Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 271–311). MIT Press, Cambridge.
- Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar* (Technical Report No. 2). Center for Cognitive Science, Rutgers University.
- Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science*, 275, 1604–1610.
- Prince, E. F. (1986). On the syntactic marking of presupposed open propositions. In A. Farley, P. Farley, & K.-E. McCullough (Eds.), *Papers from the 22nd Meeting of the Chicago Linguistic Society* (Vol. 2: The Parasession on Pragmatics and Grammatical Theory, pp. 208–222). Chicago.
- Reinhart, T. (1982). Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, 27(1), 53–94.
- Rochemont, M. S. (1986). *Focus in generative grammar*. Amsterdam: John Benjamins.
- Samek-Lodovici, V. (1996). *Constraints on subjects: An optimality theoretic analysis*. Unpublished doctoral dissertation, Rutgers University, Piscataway, NJ.
- Schneider-Zioga, P. (1994). *The syntax of clitic doubling in Modern Greek*. Unpublished doctoral dissertation, University of Southern California, Los Angeles.
- Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

- Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
- Sorace, A. (1992). *Lexical conditions on syntactic knowledge: Auxiliary selection in native and non-native grammars of Italian*. Unpublished doctoral dissertation, University of Edinburgh.
- Steedman, M. (1991). Structure and intonation. *Language*, 67(2), 260–296.
- Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: John Wiley.
- Tsimpli, M. I. (1995). Focusing in Modern Greek. In K. É. Kiss (Ed.), *Discourse configurational languages* (pp. 176–206). Oxford: Oxford University Press.
- Tsiplakou, S. (1998). *Focus in Greek: Its structure and interpretation*. Unpublished doctoral dissertation, University of London, School of Oriental and African Studies.
- Valioui, M. (1994). Anaphora, agreement, and right dislocation in Greek. *Journal of Semantics*, 11, 55–82.
- Vallduví, E. (1992). *The informational component*. New York: Garland.
- Vallduví, E. (1995). Structural properties of information packaging in Catalan. In K. É. Kiss (Ed.), *Discourse configurational languages* (pp. 122–152). Oxford: Oxford University Press.
- Vallduví, E., & Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics*, 34(3), 459–519.
- Vilkuna, M. (1995). Discourse configurationality in Finnish. In K. É. Kiss (Ed.), *Discourse configurational languages* (pp. 244–268). Oxford: Oxford University Press.
- Vion, M., & Colas, A. (1995). Contrastive marking in French dialogue: Why and how. *Journal of Psycholinguistic Research*, 24(5), 313–331.

## Appendix A. The Validity of Web-Based Studies

Experiment 1 was administered using the World-Wide Web, a method that has proved controversial in the recent experimental literature (e.g., Mehler, 1999; Johnson-Laird & Savary, 1999). It has been argued that by using web data, the experimenter can exercise less control over the experimental setting, as each subject might complete the experiment under different conditions, possibly in an environment that includes noise or other distractions. Also, there is an obvious need for making sure that the subjects taking part in the experiment respond in the way intended by the experimenter, i.e., that they understand and follow the experimental instructions properly. A third problem is subject authentication—we have to guarantee that the subject provides genuine data and does not take part more than once in each experiment.

In this Appendix, we will discuss how our experimental software is designed to address these problems. Also, we will present the results of a validation study that directly compares the results obtained in Experiment 1 (web-based) and Experiment 2 (lab-based).

### *A.1. Experimental Procedure*

Experiment 1 was administered using WebExp (Keller et al., 1998), a software package designed for conducting psycholinguistic studies over the web.<sup>12</sup>

WebExp is implemented as a set of Java classes. As Java is a full-fledged programming language, it gives the web designer maximum control over the interactive features of a web site. WebExp makes use of this flexibility to keep the experimental procedure as constant as possible across subjects. An important aspect is that the sequence in which the experimental items are administered is fixed for each subject: the subject does not have the possibility to go back to previous stimuli and inspect or change previous responses. (If the subject hits the ‘back’ button on the browser, the experiment will terminate.)

Another important feature is that WebExp provides timings of subject responses by measuring onset time and completion time for each response. The studies reported in the present paper make no direct use of these timings, as they only deal with acceptability judgments. Nevertheless the timings are useful to screen the responses for anomalies, i.e., to eliminate the subjects who responded too quickly (and thus probably did not complete the experiment in a serious fashion), or those who responded too slowly (and thus were probably distracted while doing the experiment). WebExp automatically tests the response timings against upper and lower limits provided by the experimenter and excludes subjects whose timings are anomalous. Further manual checks can be carried out on the response timings later on.

### *A.2. Subject Authentication*

Apart from providing response timing, WebExp also offers a set of safeguards that are meant to ensure the authenticity of the subjects taking part, and exclude subjects from participating more than once.

1. **Email address:** Each subject has to provide their email address. An automatic plausibility check is conducted on the address to ensure that it is syntactically valid. If the address is valid, then WebExp automatically sends an email to this address (containing a message thanking the subject

<sup>12</sup>The WebExp software package is freely available for non-commercial purposes. For details, please consult the site [http://www.hcrc.ed.ac.uk/web\\_exp/](http://www.hcrc.ed.ac.uk/web_exp/). A central entry page for all experiments using WebExp can be found at <http://surf.to/experiments/>.

for taking part). If the email bounces, the experimenter should exclude this subject from the data set, as they probably used a fake identity.

2. **Personal data:** Before being allowed to start the experiment, each subject has to fill in a short questionnaire supplying name, age, sex, handedness, and language background. These data allow manual plausibility checks to be conducted, and subjects that give implausible answers can be eliminated from the data set.

3. **Responses:** A manual inspection of the responses allows us to detect subjects that have misunderstood the instructions and responded in an anomalous fashion, e.g., by giving the same response to every item.

4. **Connection data:** The software also logs the following data related to the subject's web connection: internet address of the host machine, the type of computer and operating system, and the web browser the subject is using. This information (in addition to the email address) is valuable in detecting subjects that take part more than once.

Note that taking part in a WebExp study requires a subject to give up their anonymity and supply name and email address. This is a move we consider justified in the interest of ensuring subject authenticity. The experimental web site contains a privacy statement that guarantees that all subject data will be treated strictly confidential.

### *A.3. Comparison with Laboratory Data*

The safeguards outlined in the last section go some way towards ensuring the authenticity of web data by eliminating subjects that are not genuine, that take part more than once, or that misunderstand the instructions. However, the ultimate test of web-based data is a comparison with data obtained using a conventional, lab-based procedure.

To provide such a comparison, a set of statistical tests were carried out on the data that overlap between Experiment 1, which was conducted over the web using WebExp, and Experiment 2, which was conducted in the laboratory (using the same software). (The descriptive statistics for both experiments are tabled in Appendix C.) The overlapping data for both experiments constitute acceptability judgments for three different word orders (SVO, OVS, VSO) in five contexts (null, all focus, S focus, O focus, V focus), which yields a factorial design of  $ORD \times CON = 3 \times 5$  (see Sections 3.3 and 4.3 for details of the experimental design and procedure). For Experiment 2, all analyses were carried out on the average of the judgments for both accent patterns (S and O accent), in order to make these data comparable to the judgments for written stimuli in Experiment 1. (The underlying assumption is that subjects assign an accent at random if they see written stimuli, clearly an idealization.)

The safeguards listed in Section A.2 were applied consistently across both experiments. In Experiment 1, the data of two out of 40 subjects were discarded after an inspection of the responses showed that they had not completed the task adequately (5% of the sample). The percentage was comparable in Experiment 2, where the data of two out of 36 subjects were excluded (6% of the sample).

#### *A.3.1. Predictions*

Our hypothesis is that there is no difference between the response patterns obtained over the web and in a laboratory setting. If this hypothesis is correct, then the same significant effects should be obtained for both data sets. Furthermore, we can perform an ANOVA on the combined data from the web-based and the lab-based studies, treating the experimental procedure as a between-subjects

factor. Under the hypothesis that there is no difference between the two data sets, a main effect of the experimental procedure, and in particular, interactions between the procedure and the other experimental factors should be absent. Finally, we can test the hypothesis that there is a linear relationship between the web-based and the lab-based judgments by performing a correlation analysis on the two data sets.

### A.3.2. Null Context Condition

As in Experiments 1 and 2, the non-context condition and the context condition were analyzed separately. In the non-context condition, the ANOVA for the web-based study yielded an effect of order, which was significant by subjects only ( $F_1(2, 32) = 14.55, p < .0005$ ;  $F_2(2, 14) = 3.18, p = .156$ ). An effect of ORD was found also in the lab-based study, this time significant both by items and by subjects ( $F_1(2, 32) = 22.51, p < .0005$ ;  $F_2(2, 14) = 37.29, p < .0005$ ). A further ANOVA was carried out on the combined data from the web-based and the lab-based study, treating the experimental procedure EXP (web or lab) as a between-subjects variable. Again, we found a highly significant effect of word order ( $F_1(2, 54) = 30.53, p < .0005$ ;  $F_2(2, 14) = 7.78, p = .005$ ). There was a main effect of experimental procedure, which was significant by items, but non-significant by subjects ( $F_1(2, 54) = 2.16, p = .154$ ;  $F_2(1, 7) = 123.10, p < .0005$ ). That EXP was significant by items has an obvious explanation in the fact that Experiment 1 used written materials, while Experiment 2 used speech stimuli. As predicted, there was no interaction between EXP and word order, indicating that the type of experiment (web or lab) did not interfere with the overall word order effect.

### A.3.3. Context Condition

The ANOVAs for the context condition confirmed the results for the null context. On the web-based data, we found significant effects of ORD ( $F_1(2, 32) = 34.68, p < .0005$ ;  $F_2(2, 14) = 12.25, p = .001$ ) and CON ( $F_1(3, 48) = 3.37, p = .026$ ;  $F_2(3, 21) = 5.46, p = .006$ ). An interaction of ORD and CON was also present ( $F_1(6, 96) = 2.42, p = .002$ ;  $F_2(6, 42) = 2.42, p = .043$ ). As predicted, the lab-based study showed exactly the same pattern: there was a significant effect of ORD ( $F_1(2, 32) = 19.93, p < .0005$ ;  $F_2(2, 14) = 9.49, p = .002$ ) and CON ( $F_1(3, 48) = 3.98, p = .013$ ;  $F_2(3, 21) = 6.10, p = .004$ ), and an interaction of the two factors ( $F_1(6, 96) = 7.06, p < .0005$ ;  $F_2(6, 42) = 4.06, p = .003$ ).

We also carried out an ANOVA on the pooled data from the web-based and the lab-based study, using the experimental procedure EXP (web or lab) as a between-subjects variable. Again, the effect of ORD, CON, and the interaction ORD/CON were highly significant. As in the null context condition, there was a main effect of EXP, which was significant only by items, ( $F_1(1, 32) = .81, p = .375$ ;  $F_2(1, 7) = 12.33, p = .010$ ). Again, this is probably an effect of stimulus type (written or speech). As predicted, there was no interaction of EXP with and any of the other factors, indicating that the experimental procedure did not affect the overall acceptability pattern.

Finally, we conducted a correlation analysis that compared the judgments for each cell in the web-based and the lab-based data set. For the context condition, a highly significant correlation was obtained both by subjects ( $r(12) = .895, p < .0005$ ) and by items ( $r(12) = .917, p < .0005$ ). This can be considered strong evidence that the subjects behaved in a similar fashion under both experimental conditions. (A correlation analysis for the null context condition could not be conducted, as the number of data points was too small.)

*A.4. Discussion*

We have outlined the safeguards that a sophisticated software package such as WebExp puts in place for subject authentication in web-based experimental studies. We then presented a re-analysis of the data obtained in Experiments 1 and 2 to back up the claim that web-based experimental data and laboratory data yield comparable results. Separate ANOVAs on both data sets revealed the same significant main effects and interactions. We also failed to find by-subjects effects of experimental procedure in an ANOVA on the pooled data, both for the null context and for the context condition. By-item effects of the experimental procedure were obtained, which can be explained straightforwardly by the fact that the web experiment used written stimuli, while the lab study employed spoken materials. Crucially, there was no interaction between experimental procedure and the other experimental factors. We further showed that there is a high correlation between the average judgments obtained with both procedures. Taken together, these results suggest that there is no difference between web-based and lab-based data, at least as far as the purpose of the present paper is concerned. (For further validation studies dealing with web-based psycholinguistic data, the reader is referred to Corley, Keller, & Scheepers, 2000.)



## Appendix B. Materials

The following is a list of the lexicalizations used in Experiments 1 and 2. Only the SVO condition is listed; the other word orders can be derived by permutation in accordance with example (22). The materials with clitic doubling used in Experiment 2 can be derived by adding a clitic in front of the verb, as illustrated in example (25).

- (32) O Petros      tha plini      to aftokinito.  
the Petros-NOM will wash-3SG the car-ACC  
'Petros will wash the car.'
- (33) O Kostas      tha stili      tin prosklisi.  
the Kostas-NOM will send-3SG the invitation-ACC  
'Kostas will send the invitation.'
- (34) O Tasos      tha diavasi   tin efimerida.  
the Tasos-NOM will read-3SG the newspaper-ACC  
'Tasos will read the newspaper.'
- (35) O Giorgos    tha grapsi    to senario.  
the Giorgos-NOM will write-3SG the script-ACC  
'Giorgos will write the script.'
- (36) I Eleni      tha pulisi    tin afisa.  
the Eleni-NOM will sell-3SG the poster-ACC  
'Eleni will sell the poster.'
- (37) I Christina   tha agorasi   to pagoto.  
the Christina-NOM will buy-3SG the ice-cream-ACC  
'Christina will buy the ice-cream.'
- (38) I Maria      tha di        to spiti.  
the Maria-NOM will see-3SG the house-ACC  
'Maria will see the house.'
- (39) I Ana        tha pari      to leoforio.  
the Ana-NOM will take-3SG the bus-ACC  
'Ana will take the bus.'

What follows is the list of lexicalizations used for the contexts in Experiments 1 and 2. Only the materials for the S focus context are listed; the other contexts can be derived in accordance with example (23).

- (40) Pios tha plinei to aftokinito?  
'Who will wash the car?'
- (41) Pios tha stili tin proskisi?  
'Who will send the invitation?'
- (42) Pios tha diavasi tin efimerida?  
'Who will read the newspaper?'
- (43) Pios tha grapsi to senario?  
'Who will write the script?'
- (44) Pios tha pulisi tin afisa?  
'Who will sell the poster?'
- (45) Pios tha agorasei to pagoto?  
'Who will buy the ice-cream?'

- (46) Pios tha di to spiti?  
 ‘Who will see the house?’
- (47) Pios tha pari to leoforio?  
 ‘Who will take the bus?’

The question-answer pairs in (48) and (49) were used as modulus items for Experiments 1 and 2, respectively. In both cases, the answer sentence is not fully grammatical. In (48), the numeral *dodeka* should be preceded by the prepositional phrase *stis* instead of the definite article *tin*. In (49), the adverb *puthena* ‘nowhere’ should follow the verb, instead of preceding it.

- (48) Ti ora girise o Nikos htes? O Nikos girise tin dodeka.  
 which hour returned-3SG the Nikos-NOM yesterday the Nikos-NOM returned-3SG at twelve  
 ‘When did Nikos return yesterday? Nikos returned at twelve.’
- (49) Pou tha tin pai tin Eleni o Vasilis? Den tha PUTHENA pai o  
 Where will her take-3SG the Eleni-ACC the Vasilis-NOM? Not will nowhere take-3SG the  
 Vasilis tin Eleni.  
 Vasilis-NOM the Eleni-ACC.  
 ‘Where will Vasilis take Eleni? Vasilis won’t take Eleni anywhere.’

## Appendix C. Descriptive Statistics

Table 12: Descriptives for Experiment 1, Null Context Condition

word order	mean	SD	SE
SVO	.5660	.3588	.0870
OVS	.4285	.3434	.0833
VSO	.4830	.3614	.0877
VOS	.4679	.3487	.0846
SOV	.3637	.3112	.0755
OSV	.3455	.3387	.0822

Table 13: Descriptives for Experiment 1, Context Condition

context	word order	mean	SD	SE
all focus	SVO	.3817	.3312	.0803
	OVS	.0673	.3556	.0863
	VSO	.2658	.2401	.0582
	VOS	.2362	.2680	.0650
	SOV	.1185	.2969	.0720
	OSV	.0651	.2925	.0709
S focus	SVO	.3619	.2802	.0680
	OVS	.2526	.2637	.0640
	VSO	.2568	.2978	.0722
	VOS	.1933	.3304	.0801
	SOV	.0738	.2149	.0521
	OSV	.1225	.2490	.0604
O focus	SVO	.3626	.2870	.0696
	OVS	.2299	.2694	.0653
	VSO	.1484	.2652	.0643
	VOS	.2407	.2087	.0506
	SOV	.1583	.3024	.0733
	OSV	.1320	.2769	.0672
V focus	SVO	.3168	.3194	.0775
	OVS	.0877	.3400	.0825
	VSO	.1311	.2681	.0650
	VOS	.2181	.2780	.0674
	SOV	.1160	.2299	.0558
	OSV	.0383	.1972	.0478

Table 14: Descriptives for Experiment 2, Null Context Condition

word order	mean	SD	SE
Svo	.3475	.1991	.0575
svO	.4373	.1708	.0493
ovS	.2604	.2204	.0636
Ovs	.2777	.1751	.0505
vSo	.3096	.2224	.0642
vsO	.3441	.1909	.0551
ScIvo	.3524	.1888	.0545
sclvO	.1943	.2006	.0579
oclvs	.3707	.1875	.0541
Oclvs	.1728	.2097	.0605
clvSo	.3453	.2230	.0644
clvsO	.2323	.2004	.0579

Table 15: Descriptives for Experiment 2, Context Condition

context	word order	mean	SD	SE
all focus	Svo	.3506	.2732	.0663
	svO	.4112	.2278	.0552
	ovS	.1892	.2900	.0703
	Ovs	.1728	.3941	.0956
	vSo	.3318	.3547	.0860
	vsO	.4254	.2053	.0498
	ScvO	.2754	.4216	.1023
	sclvO	.2896	.2565	.0622
	oclvs	.3912	.2006	.0487
	Oclvs	.1119	.3754	.0910
	clvSo	.3720	.2577	.0625
	clvsO	.2802	.2572	.0624
	S focus	Svo	.5261	.2412
svO		.3790	.2408	.0584
ovS		.4077	.1811	.0439
Ovs		.1474	.2251	.0546
vSo		.4096	.2083	.0505
vsO		.2756	.2419	.0587
ScvO		.4972	.2359	.0572
sclvO		.2685	.2280	.0553
oclvs		.4892	.2651	.0643
Oclvs		.1758	.2659	.0645
clvSo		.5155	.2105	.0510
clvsO		.2695	.2330	.0565

Table 16: Descriptives for Experiment 2, Context Condition (Continued)

context	word order	mean	SD	SE
O focus	Svo	.2015	.3553	.0862
	svO	.5391	.2091	.0507
	ovS	.2716	.2733	.0663
	Ovs	.4665	.3926	.0952
	vSo	.1181	.4128	.1001
	vsO	.3307	.2316	.0562
	ScIvo	.0797	.3289	.0798
	sclvO	.1156	.3251	.0788
	oclvS	.2285	.2669	.0647
	Oclvs	.2513	.3029	.0735
	clvSo	-.0165	.3052	.0740
	clvsO	.1180	.2705	.0656
	V focus	Svo	.2682	.2511
svO		.4039	.2962	.0718
ovS		.1369	.3374	.0818
Ovs		.1813	.2559	.0621
vSo		.2083	.2576	.0625
vsO		.3191	.2731	.0662
ScIvo		.1789	.3775	.0916
sclvO		.2836	.2358	.0572
oclvS		.1811	.4277	.1037
Oclvs		.0833	.3596	.0872
clvSo		.2782	.2607	.0632
clvsO		.1618	.4067	.0986