Preliminary Minutes - UTC #74 & L2 #171
Mountain View, CA - December 5, 1997

1. ADMINISTRATIVE ISSUES

1A. UTC Membership Roll Call:
PRESENT: Apple Computer, Inc.; Digital Equipment Corporation; Hewlett-Packard Company; IBM Corporation; JustSystem Corporation; Microsoft Corporation; Novell, Inc.; NCR Corporation; Oracle Corporation; The Research Libraries Group, Inc.; Sun Microsystems, Inc.; Sybase, Inc.; Unisys Corporation.
(Total present: 14)

NOT PRESENT: Booz, Allen & Hamilton, Inc.; Reuters, Ltd.; Silicon Graphics;

LIST OF ATTENDEES: See attached appendix

1B. Declaration of Joint Meeting

1C. Approval of Joint meeting Agenda [97-261R]
[#74-M1] Motion: To approve agenda as amended.
Moved by Jenkins, seconded by Carroll.
Unanimously Approved

1D. Approval of Minutes and Review of Action Items [97-255]
[#74-M2] Motion: To approve minutes of previous meeting as amended.
Moved by Jenkins, seconded by Sargent
Unanimously Approved
Action Items: see attachment for details

1E. Registration of new documents

1F. Review of Meeting Calendar (see attachment)
Action 74-1 for Aliprand: Place L2/97-262 on Agenda for Next Meeting (from Uma).

Action 74-4 for Aliprand, Winkler: Verify availability of quorum for April meeting. Important for Feb/Apr meeting to prepare/react to SC2 meeting - vital regarding Version 3.0.

2. TECHNICAL ISSUES

2.A. ERRATA TO UNICODE 2.0
2.B. PROPERTIES -- HANGUL SYLLABLE DECOMPOSITIONS -- CANONICAL OR NOT? (Davis/Edberg)

Davis: Hangul decomposition. Canonical decomposition or compatibility. Question whether we take action or not.

Whistler: Part of conformance chapter that spells out in detail Hangul conjoining characters. Not a connection of how formal. (2.6)

Davis: We don't have name charts. Can be algorithmically decomposed.

Freytag: In Chapter 3, one aspect of decomposition -- are you forced to make the distinction if you haven't changed the data?

Jennings: Issue of decomposing or composing when we use Hangul.

Sargent: What are the advantages of decomposing? And, what do the Koreans say?

Davis: If you recognize Hangul, you need to have equivalence

Freytag: Implying a preference gets you into trouble.

McGowan: If there is a direction, greater control if you decompose.

Davis: There is more expressive power in decomposition.

Uma: 1. Are they equivalent? And, 2) which form do you prefer?

McGowan: Compatibility as compared to canonical.

Freytag: In the process of writing 3.0, "equivalent" means equivalent both ways. Dispersed throughout the book is preference to decompose. Hangul will clarify.

Davis: In Section 3.10, on p. 3-12 Hangul the phrase "canonical syllable block" is used. Anyone who looks at it, would read as canonical.

McGowan: ... support of code to do composition or decomposition.

Hangul syllable decomposition
[#74-M3] Motion: In section D23, canonical decomposition
(page 3-7), change the final "." to "," or by application
of the combining algorithm defined in section 3.10."
Moved by Whistler, seconded by Davis
Unanimously Approved

Action 74-5 for Editorial Committee: Re decomposition: revise text & tables, as necessary to make it clear that canonical equivalence.

Proposal for Version 2.1 [97-253]

Whistler: Minor changes that we are making like this may be rolled into 2.1 document.

Freytag: Putting something in 2.1 really highlights, so look at in that context.

Davis: It definitely belongs in 2.1 - clarification in 2.1. Look for changes to be made in text later for 3.0.

2.C. REVISED PROPOSALS

2.C.1.   **Plane 14 Characters for Language Tags (Whistler) [97-171R2]**
2.C.2.   **Comments on Plane 14 Position Paper (Mati Allouche) [97-256]**
2.C.3.   **Publish Plane 14 Proposal as a UTR? (Freytag)**

Uma: Important to make fit format for WG2. As soon as Glenn Adams is finished with RTF, make proposal for WG2.

Suignard: Propose this as a New Work Item for WG2. Deadline for WG2 documents is Dec 12.

Uma: What is the reaction from IETF side?

Freytag: There is no technical difference between September 18th version and Glenn's IETF draft. Put a header on the September draft and make it a UTR. Then submit as contribution to WG2.

Uma: This notion of tagging has not been discussed in WG2. We don't have to push as a proposal.

Freytag: But it is demanded by an important member of the implementer community.

Suignard: [volunteering to help with a proposal to WG2]. If we can get 1) to take a resolution to WG to approve, 2) we can get SC2 to present a proposal for project.

Freytag: Submit paper work. Technical paper gets mailed in advance.

Action 74-6 for Whistler: Fix BNF error. Put in UTR #7. Make Plane 14 language tagging proposal into Unicode Technical Report #7.

Action 74-7 for Whistler: Send UTR #7 to Mike Ksar for distribution with the December WG2 with cover memo "For consideration by WG2".
Action 74-8 for Suignard: Prepare a draft CD based on UTR #7 for WG2 discussion.
Action 74-9 for Suignard: Prepare application to SC2 for NP subdivision, to allow Plane 14 use in ISO/IEC 10646.

2.D.1. PROPOSAL FOR VERSION 2.1 (Moore for Editorial Committee) [97-253]

Proposal for Version 2.1 as UTR #8, using document 97-253 as the basis.

Moore: Everyone has heard this discussed. We wanted to take a combination of small or important changes: 1) U+FFFC Object Replacement Character; and, 2) Euro sign, which is in the process of being accepted. We restricted (2.1) to a small set.

Jenkins: (in response to a request from the floor to correct some glyph errata) I would prefer not to correct glyph errors in 2.1. 1) we don't have font; 2) if we do the glyph errors, we should do all of them and we aren't prepared to do so.

Moore: My initial approach is to not include glyphs.

Uma: what about "corrections to be done"? Go ahead with all known corrections.

Becker: Document 253 needs a header to identify what it is. And a date.

Davis: On glyphs... we should agree that this is a good thing.

Uma: We are not far from 3.0. Should we go ahead with this?

Ksar: I don't understand 2.1. It is very similar to 1.1 (as a TR) Cannot call as Version 2.1. What is the timeline for 2.1 and the timeline for 3.0?

Whistler: March 1999. 3.0 is a complete republication, with 10,000 new characters. Freeze date is synched with March WG2 meeting. In March 1998 we will freeze things likely to go into 3.0. Year set back on publication. Prospective date on 2.1 is today.

Ksar: Freeze date of March 1998? If you are going to do vertical extension, you can't freeze that early.

Davis: Goal should be something to refer to as a "conformance standard". We are finding people are needing to refer to (for example) updates to BiDi and have to have referable "point" versions as part of the standard. Companies can't wait that long for published versions. We need to make it

clear that this is an addendum, a set of modifications to Version 2.0. I thought it would include contents of Errata on Web site. All those that have been agreed upon by UTC.

Texin: Good collection and agree with intent. But concerned about publishing too quickly. Too many others are relying on Unicode. Some companies still working with 1.1. Need to communicate that releases are coming, need to make business decisions.

Freytag: Object Replacement Character is being used by many companies for years. Being forced into a situation of non-conformance is worse than not being forewarned. The Euro is needed. Not designed -- and we should stress -- to delink from 10646, only ones that have high chance of being approved for 10646 are in. It will formalize informal business practice.

It is incorrect to say it is not Version 2.1. You need a place to conformant with. A document is not a book. We tend to publish our delta documents as UTRs. Not an update - it is Version 2.1

Whistler: Notice and frequency. (directed to Texin) If this is the last delta before 3.0, I will be very happy.

Jenkins: Being driven to get the Euro in 3.0 is not soon enough. Nice to add corrections.

Becker: Is there anything we need to or can do for blessings from SC.?

Ksar: I don't think this is anything. It's in the bucket. If anything, for the Euro. Reason is that it is important. There are many discussion. Many say Unicode is not ready.

Aliprand: Publicity mileage that Unicode supports the Euro.

Davis: Put this out when necessary; put corrigenda at same times and get it out. Goal should be to put out updates where it makes sense, without eating into 3.0 effort. Give a way to be conformant.

[#74-M4] Motion: That the UTICA approves production of
Unicode Technical Report which describes Version 2.1,
using document 97-253 as the basis of the UTTER.
Moved by Davis, seconded by Honomichl
Unanimously Approved

Davis: As far as content, include all errata currently on web site; e.g. decomposition with no break space.

Freytag: I second Mark's request for errata. I would like to see the Euro sign description split between history and technical issues.

Jenkins: Regarding characters beyond Euro and OBI. Ethiopia is the only approved script. This does not add much for us. Slow down editing of this document. We don't have fonts. Technically more difficult to get this document out, not insurmountable, but very slow. Include list of scripts that are under consideration.

Moore: I second what John said. I don't want to hold things up. We were delayed because fonts were not available.

Suignard: Include small additions to exiting scripts. Romanian for example. Good gesture.

Whistler: 1) I am seriously opposed to adding any more characters to 2.1. These two are obvious; and don't go beyond that. If six months down the road someone is getting beat up business wise - then do 2.2. 2) I am opposed to adding glyphs to this document. Errata to do this. 3) Notion that we should include "the errata" is not well defined.

Freytag: Ken makes some good points. A possible compromise: definite list and links to web location. Make it possible for a payer mailing. Satisfies the need for those who say glyphs are wrong. We acknowledge that we know that glyphs are wrong. Have 2.1 out the door by February. Compile list of which errata are to be included. We have an urgency. The best way to sink a fast boat is to overload it.

Ksar: I agree with Asmus - to have an errata glyph list. Mention that characters approved by SC2 and WG2. Include the list, information in 3.0. This addresses Tex's point.

Davis: Good suggestion. Minor errata, glyphs and all approved characters on the web with URL links. Cut this off, put it out.

Whistler: Additional characters - 9000+ - major data reconciliation problem to identify them. Better to point to pipeline.html. Don't publish

Uma: Pointers should be there to the pipeline. Define 2.1 conformance.

Freytag: Informative links to future links - 1) Extremely explicit links to errata... only those published in 2.1; or, 2) one link to list of errata. Stuff on web which is variably approved. "These approved as of date..."

Moore: I'm happy to have a list of all approved errata.

Davis: Everything that I put on the web site was approved as errata by the UTC. What is not?

Becker: BiDi must be packaged in this document.

Uma: Errata needs to be examined for acceptance/non-acceptance by UTC.

Jenkins: Glyph errata have not been approved by the UTC. Not an urgent "need" to have glyph list.

Aliprand: Is there anything in 97-253 that people disagree with?

Davis: p. 14 "Zero width no-break space" is not a space. We have breaker and spaces. Compatibility mapping of ZWNB space. Should not map to a space.

McGowan: This functions as byte order mark on my system!

Moore agreed to take this out.

Sargent: Mathematical character property. Original idea was to mark a character as a math property indicator, then presumably will be used with mathematics.

Davis: Makes algorithms easier and clearer. Mapping compatibility characters is determined by what it maps to. Easier to say these are mathematical symbols.

Sargent: comfortable with that?

Aliprand: More information about the Euro is needed. We have to make it clear which character is the Euro.

Freytag: We need to include a Euro glyph on our site.

Whistler: This (2.1) will be on html.

Whistler: changes on p. 3; p. 10; p. 11 FFFD --> FFFC; p.12 Raise question re decompositions... problem of middle dot.

Davis: Middle dot used by millions of people

Whistler: Catalan letters are not decomposable.

Davis: Include informative note re modifier letter middle dot.

Freytag: The light dot is used phrase initially, so is not decomposed. B7 is in Latin 1 and typical mapping is as in current decomposition.

Davis: You're saying, "Leave decomp. alone?

Freytag: Yes.

Davis: As an alternative, we could say - middle dot is ambiguous, like hyphen-minus. When used as light bullet it is usually at start, and set off by white space. If after an "l", likely to be an extender in European languages.

Freytag: It is correct to treat it as having word extension property, but also as punctuation.

Whistler: 00B7 should have both punctuation and extension property.

Davis: Explaining in 2.1?

Whistler: No

Freytag: We agree -->00B7 1) punctuation; 2) extender

Whistler: If 00B7 middle dot can be used as extender, then there is no need for mod. Letter middle dot.

Reversal of Motion #70-9:
[#74-M5] Motion: That the UTC rescinds acceptance of
the character U+02DF, MODIFIER LETTER MIDDLE DOT.
Moved by Davis, seconded by Whistler
Approved 9,0,2

Action 74-10 for Moore: Draft Unicode Technical Report #8 (Version 2.1) using L2/97-253 as a basis
Action 74-11 for Aliprand/Winkler: Put UTR #8 on agenda for February UTC/L2 joint meeting.
Action 74-12 for Jenkins: Revise entry for U+02DF, MODIFIER LETTER MIDDLE DOT, to show that it has been removed.
Action 74-13 for Moore: Revise properties for U+00B7, to add use as extender (in addition to punctuation)
Action 74-14 for Moore: Distribute revised properties for U+00B7 by e-mail to "unicore"

2.D.3. REPORT ON CHARACTER-GLYPH MODEL

Hart: Character Glyph model. ISO ballot passed. Comments from US and Canada. Editors have produced the final text. Does anyone volunteer to look at the final text? (5 volunteers came forth)

The Chair thanked Ed Hart for his work on the Character/Glyph Model, and congratulated him and Al Griffee on its publication.

MEETING of December 3, 1998 ADJOURNED 5 PM

MEETING of December 4, 1998 OPENED at 9 am Thursday, March 4, 1998
AGENDA ITEM:

2.D.4. Variant Tag Mechanism (Hiura/Kobayshi) [L2/97-260]

Hirura: Extension to Plane 14 extension mechanism to address ideographic variant problem. This is not necessarily bound to locale (as current language Plane 14 proposal is). If we do not address variant problem, use of user defined characters affects inter-operability. The proposal says to use Base character as reference and add a reference number with single shift modification mechanism (rather than locking shift). Processing system can strip it out or skip.

Becker: Do you propose to standardize explicit variants?

Hiura: Yes. How far we want to cover categories is open to discussion.

Suignard: I want to be cautious. Have to define new equivalents, especially when they have been disunified. Could use it for other purposes, to disunify.

Hiura: Most new characters can be identified as variants of existing base.

Uma: Another method of encoding the same thing

Freytag: Another aspect. Gain one thing: have immediate fallback – ignore variant tag, no changed needed. Are the variants primarily visual differences, or are there meaning differences.

Kobayashi: The last example is a rare case. The ideographs are example (last one) are identical in meaning in Japanese, but different in Chinese.

Freytag: Having base character in stream might be more helpful than having a table for Plane 2 characters.

McGowan: This committee should not be involved in cataloging and identification of variants.

Becker: Referring to Z-variants? Really refers to glyph variants. Should limit it to small number.

Jenkins: Agree with intent. Variant tag should follow character. Concerned about font development. Restrict to variations that occur in single typeface. Excellent way to do this.

Whistler: "z variant tags" -- specifically name the variance. Should be fairly precisely defined. Tie closely to z-variant axis. Syntax not stated clearly enough - include explicit examples.

Texin: On conversion to other character sets... If it has a variant, do we assume that it doesn't exist in another character set? Is there a relationship between variant tags on one character and on another? Is
there any intent to include meaning to assist searching algorithm?

Hirua: No - have not thought about this. Semantics are identical in this context. Will investigate in terms of searching for specific variants. ...further unification with national standard sets.

Uma: Sometimes language dependent, e.g., last example.

Hirura: Attempted to consider this.

Uma: Possible to consider for other uses, e.g., math use. Maybe need other information as well.

Moore: Useful for large gaiji sets. Good proposal, worth pursuing.

Ksar: It is worth pursuing. I'm concerned about extension to multiple purposes.

Suignard: I would like to restrict to Han characters, with separate project on other uses.

McGowan: Analogous to language tagging proposal. Content is out of scope for the committee. This is a near higher-level protocol. Provide a raw mechanism. This is what is needed. Someone else should be responsible for the details.

Uma: Replacement for general variation mark?

Jenkins: This is an in-between one.

Whistler: Specific comments. Tags be called Z-variant tags. Be tied specifically to CJK unification model.

Texin: Variants should be defined outside of this committee. Issue of compatibility with other character sets?

Hiura: National standards are evolving. This is an alternative to hard coding in other planes. Generally use user-defined characters in national context for such variants.

Freytag: Have nice clear restriction [to CJK]. Use of user-defined characters is bad enough, no other user-defined items. I second Ken's comments. I am uncomfortable with "open" variant marks. While IRG may be best choice, this proposal should not be submitted to WG2 until we have an initial set of variants. I would not like to present mechanism without initial set of characters that are variant.

Hiura: Intent was not to define set. IRG would be group to produce set.

Action 74-15 for Kobayashi/Hiura: Do revision of Variant Tagging proposal.

3. SPECIFIC SCRIPTS
3.A. WHOLE SCRIPTS

3.A.1. REVISED PROPOSAL
a. Western Music (McGowan) [97-129R]

Aliprand: The Music Library Association made comments and Perry Roland is willing to add the four characters they suggested.

McGowan: Everson made comments. Code chart horizontally oriented. Go across - will be cleaned up for publications.

Becker: Dingbat situation? We need to consider pre-existing data.

McGowan: Dingbats include sharp, natural and flat.

Freytag: No point in disunifying.

Jenkins: Crosses plane boundry.

McGowan: Difficulty of implementing. Fatal to close-up their slots.

Freytag: I think we should not go so fast.

McGowan: Only sharp, natural and flat now used in text.

Freytag: Idea of a form to use in 2D layouts --> different story.

McGowan: Removing 3 slots with no cross references is a mistake.

Ksar: If we have 3 on Plane 0, how do we handle if we have both?

Whislter: I am open to Asmus' argument to leave out the three, but this will be the preferred standard for music implementation, if accepted.

Freytag: The three positions could be used by implementers.

Becker: Are these characters to be used for diagrammatic layout or not? We should say so. Textural uses or layout of score.

Honomicl: Two dimensional, no difference between 2D for math and 2D for music. Rick, is this harder to sell?

Aliprand: You are going to get things other than sharps, flats, naturals in text. There needs to be only one set.

Ksar: Since these will be going into Plane 1, idea of having 3 more code locations in Plane 1 is not as excessive. I want to make sure if there is a different use, then we should have those characters in Plane 1. If not, then leave them out.

Whistler: 1) the conservative position --> leave three holes; can add back in later. 2) proposal states: set to be used for higher level protocols.

McGowan; I know it works!

Bishop: 1) do the three symbols occur in other standards? Yes --> Z39.47
2) There is a lot of real estate. It seems like we ought to do it.

Jenkins: Experience has shown it is enormously confusing to have two things that are very much alike. Leave the holes. Why not quarter notes and eighth notes?

McGowan: These are legacy data of dingbats.

O'Donnell: Very confusing that they look the same.

McGowan: Used in the production of scores.

Hart: Concerned with round trip integrity

Becker: What are the bi-directional implications of these?

McGowan: All western music is left to right.

Becker: State in proposal. Explain use of control characters for slur.

McGowan: Arabic uses completely different musical tradition.

Becker: How does a phrase break around a line?

McGowan: Beaming always terminates. Phrases and slurs continue and wrap to the left. It is well defined.

Suignard: As an implementer. I hate oversimplification. There is a lot of context. Preserve ambiguity - those characters can be used for lots of purposes. I prefer to have those three disunified.

Freytag: Sharp, flat, eighth note do appear in two or more sets.

Aliprand: Sharp and flat are in library set.

Freytag: Michel raised important point. We could insist on a second encoding of the three but I'm not sure that is the right thing. At least for a couple of years we have to map to legacy data, there are enough platforms to force you to map into another character set to show the things (a five year problem). We have to be careful. This is not an easy decision. Michel said we have unambiguous code....

Whistler: Seriously missing from this proposal: 1) explicit discussion re rational for unification or disunification (why dot or staccato?) 2) as well as completion. We need to be explicit.

Sargent: Complaints of over unification from Microsoft. Makes sense to use in 2D layout.

Freytag: Please don't wait until time to put in the book to write the necessary script. I think that we should do what is necessary for this proposal.

Aliprand: People who see a problem, start a discussion.

Aliprand: Address BiDi and start end controls. Also, technical issues with discussion on this on Unicore.

Jenkins: Nice to say that this is an approved proposal, but there are two issues:1) directionality, and, 2) deciding to unify or disunify.

Uma: Leave three as holes.

WESTERN MUSIC SYMBOLS [97-129R]
[#74-M7] Motion: That the UTC accept the Western
Music Symbols proposal in principle, with the following
editorial clarifications:
* BiDi aspects of music;
* use of start and end controls;
* unification/disunification rationale (e.g., why various
dots in the proposal are discrete entities that cannot be unified);
* reservation of the positions in the proposal currently
occupied by the sharp, flat, and natural, which are
already encoded in the Miscellaneous Symbols block;
* to consider the characters U+2669 QUARTER NOTE,
U+266A EIGHTH NOTE and the beamed notes U+266B
and U+266C to be strictly Miscellaneous Symbols.
Moved by Whistler, seconded by Ksar
Approved 11,1,0

[#74-M8] **Motion:** That the UTC accept the four symbols
identified by the Music Library Association in document
L2/97-264 and request Perry Roland to include them in
the revision of the Western Music Symbols proposal.
Moved by McGowan, seconded by Whistler
Approved 11,2,0

Action 74-16 for McGowan: Revise Music proposal:
Work with Perry Roland on revision of Western Music Symbols proposal.
a) Incorporate changes requested as acceptance (BIDI aspects, etc.), b) Add new characters
proposed by Music Library Association and accepted by the UTC.

Action 74-17 for Jenkins: Add acceptance of characters proposed by Music Library Association to
the Unicode Web site.,

3.A. WHOLE SCRIPTS
3.A.2. Khmer

McGowan: Small group discussing on Unicore: Maurice Bond, Glenn Adams, etc. We have a
violent clash of experts. UTC agreed to accept Khmer repertoire a couple months ago. Glenn
revisited the decision. Two camps: 1) separate set of subjoined consonants; 2) use virama model

Aliprand: Glenn thinks that we should revoke previous decision.

McGowan: Khmer script is not used to make IndoEuropean words. Monosyllabic, Sinitic. Glenn
prefers sub-joined iconographies. Glenn has implemented both ways.

McGowan: We have deeply entrenched experts.

Bishop: There are obvious differences, but I didn't feel people were wedded to their position.
Microsoft favored sub-joined but could live with other way. Worst situation is that we don't take
advantage of Unicode and resolve.

McGowan: If we don't encode Khmer, Mongol, and Burmese in Version 3.0 we are a pathetic
group.

Aliprand: Someone needs to write new proposals.

Bishop: I will do it. One caveat. I need to make sure I can do this. Contingent upon management.

Action 74-23 for Bishop: Coordinate change in lead for Khmer and Burmese with Adams.
Action 74-24 for McGowan: Add Khmer list people to Unicore DL as needed
Action 74-25 for McGowan: Send latest copies of proposal for Khmer to Bishop
Action 74-26 McGowan: Send latest copies of proposal for Burmese to Bishop

McGowan: The highest priority is to get into Version 3.0. This has been on the table since 1988!
I want one model.

The two camps are: 1) Virama - Lee Collins, Michael Everson, Maurice Bond; 2) Sub-joined
consonant - Martin Durst, Glenn Adams, Avery Bishop

Khmer
[#74-M6] **Motion:** That the UTC officially rescinds
acceptance of the proposal for Khmer script
[Motion #73-M13], and puts development of a proposal

for Khmer on hold, pending resolution of the sub-joined
consonants vs. virama issue.
Moved by Jenkins, seconded by Freytag
Unanimously Approved

Ksar: Will we have a proposal for next UTC meeting in February? For March WG3?

Uma: Shoot for getting proposal for March WG3 meeting.

Bishop: We should move the discussion to Unicore.

Freytag: I agree. The "voters" will need to be persuaded, even though we aren't experts -
especially with arcane scripts. We need to present the evidence.

Action 74-27 for Bishop: Request management approval to work on Khmer and Burmese.

3.A. SPECIFIC SCRIPTS - WHOLE SCRIPTS
3.A.1. NEW PROPOSALS
a. Additional Canadian Syllabics (Vermeulen/Everson) [97-274] CASIC examined missing
characters identified by implementers in Canada.

Additional Canadian Syllabics
[#74-M9] Motion: That the UTC accepts the additional
Canadian syllabics missing from DAM 11 [97-274].
Moved by Umamaheswaran, seconded by Jenkins.
Unanimously Approved 13, 0, 0

3.B. SPECIFIC SCRIPTS -- IDEOGRAPHIC ADDITIONS
3.B.1. Existing compatibility characters also in CJK Extension A (Moore)
[97-254]

Moore: 254R and 279 Discovery of two characters. Recommendation to IRG to remove, which
they decided not to accept. IRG took position of it being ok to duplicate. IBM would like to have
only one mapping.

Jenkins: Clarify in discussion. IRG does not own "how" a character is encoded. They are
identifying the characters. There is no guarantee in WG2 that they take all ideographs in vertical
extension A. We define characters in Unicode. They are in the compatibility zone.

McGowan: The IBM position is that they should not be encoded a second time in Vertical
Extension A.

Jenkins: These characters are in Unicode because IBM asked for them. IRG originally didn't want
them but subsequently took them up in Vertical Extension A.

Ksar: Good process that IBM and others are going through to avoid duplicate characters.
Behooves us that they we make sure that are not included in Vertical Extension A. None of us
"experts" have checked to verify. We need to make sure that they are unique. We don't want any
duplicates.

Avery: Good to have a history but be careful not to point fingers. There seems to be an
understanding that compatibility characters are not as important. Don't duplicate.

Hiura: [Referring to document 279] Status of R Zone? We are concerned that we are redefining
the R zone - a corporate character.

Jenkins: Recommendation: repeat what was done for the music proposal and leave holes.

McGowan: We are being called upon to get in on the bickering of IRG.

Whistler: We are being asked to intervene in a dispute. I agree with Koyabashi. I don't want us to screw up. The R-Zone issue: UTC has moved away from the R-Zone. Artifact in 10646 designating characters that don't mean anything in 10646. Proceeding to intentionally duplicate Chinese characters in the vertical extension would be a mistake.

Freytag: Statement that the R-Zone is not relevant to Unicode is evidence of differences in committee. Not all distinctions in 10646 are relevant. I recommend that the UTC make recommendation to remove the two characters.

Hart: 1991 Paris meeting - Japanese representative was concerned over inclusion of compatibility.

Becker: These are there because they "snuck-in" politically. Let us declare that the 32 characters are "real" character in the standard. We agree they are in a funny place.

O'Donnell: The fact that they are in the compatibility zone, makes them second class citizens. Is there a way to change the name "compatibility."

Ksar: R-Zone was created to serve needs. We wanted to put them somewhere. It is not restricted; they were put there because they are old.

Jenkins: Vertical Extension A researched by IRG. If we feel that IBM 32 are needed. Ask IRG to do the other 30. No guarantee that Vertical A will be encoded. UTC position should be: They belong in Extension A, but we do not want them encoded again.

Koyabashi: Because of unification rules, almost all are duplicated except two. Why do they have to be treated differently?

Whistler: Two are missing. All others are known duplicates.

Koyabashi: Because they are duplicate characters, we could just ignore 30 already duplicated.

Suignard: Not a new problem for implementers.

Jenkins: These are two characters in a huge character set, two rarely used. Vendor standard has both. Not just verification.

Freytag: Round trip compatibility. Both are encoded.

Whistler: One copy now. Encoding in Vertical Extension is to duplicate extension.

McGowan: Extension A has within it two characters, not variants, already encoded on the BMP. IBM is requesting the UTC not encode these again.

Freytag: Customer impact is relatively little. Unicode has no R Zone.

Koyabashi: UTC can request WG2 to remove two characters.

McGowan: No, refrain from encoding them!

Becker: On the other hand, there are many duplicates introduced by Taiwan.

Ksar: R Zone is ISO dumping ground.

Freytag: Technical opinion - there is nothing special about these characters.

Existing compatibility characters also in CJK Extension A [97-254]
[#74-M10] Motion: To request WG2 to not encode the two
characters of CJK Unified Ideograph Vertical Extension-A,
U+4551 and U+479E, because they already exist on the BMP.
Moved by Freytag, seconded by
Approved 10, 2, 1

Request for clarification of R-Zone characters
[#74-M11] Motion: That the UTC requests WG2 to clarify
in ISO/IEC 10646 that R-Zone characters have equal status
with all other characters.
Moved by Umamaheswaran, seconded by O'Donnell.
Approved 11, 2, 0

Documentation for two compatibility characters
[#74-M12] Motion: That the UTC requests WG2 and the
Unicode Editorial Committee to document the two characters
U+FA1F and U+FA23 following the same format as for URO
characters.
Moved by Umamaheswaran, seconded by Ksar.
Approved 7, 2, 4

CJK Unified Ideographs
[#74-M13] Motion:
There are twelve characters among the CJK Compatibility
Ideographs included in Unicode 2.0 which are not
duplicates of characters among the CJK Unified Ideographs:
 U+FA0E, U+FA0F, U+FA11, U+FA13, U+FA14, U+FA1F,
U+FA21, U+FA23, U+FA24, U+FA27, U+FA28, and U+FA29.
Two of these, U+FA1F and U+FA23, are included in the
Vertical Extension A. The Unicode Technical Committee
submits the remaining ten characters to the IRG for inclusion
in its further work and requests that they be unified, mapped,
and tracked along with other proposals to the IRG.
Moved by Jenkins, seconded by Sargent
Unanimously Approved

Action 74-18 for Suignard: Prepare request to WG2 to refrain from encoding the two characters in
CJK Vertical Extension-A that are already encoded in the R-zone (as CJK compatibility
characters)
Action 74-19 for Suignard: Prepare request to WG2 to clarify the text of ISO/IEC 10646 to
indicate that R-zone characters have equal status with all other characters.
Action 74-20 for Suignard: Prepare request to WG2 that the two CJK characters in the R-zone that
are also in CJK Vertical Extension-A be documented using the same format as for URO characters
Action 74-21 for Unicode Editorial Committee: Document the two CJK compatibility that are also
in CJK Vertical Extension-A using the same format as for URO characters

Action 74-28 for Jenkins: Obtain glyphs for IBM characters in CJK compatibility block form.
Action 74-29 for Suignard: Prepare proposal to WG2 for the addition of the mirrored characters in
L2/97-250 to 10646.

3.C. ADDITIONAL NON-IDEOGRAPHIC CHARACTERS

3.C.1. Cartouche proposal for keyboard symbols (Freytag)
see below

3.A.4. SPECIFIC SCRIPTS - WHOLE SCRIPTS – NEW PROPOSALS
3.A.4.a. Disunification of Coptic

Jenkins: The rationale historically no more related (Coptic and Greek).

Davis: Is a member going to make a motion to disunify?

McGowan: Why were they put together? Give me some background.

Ksar: What are we trying to do?

Whistler: Discussed on e-mail list, most likely eight or so letters - Coptic letters. Nothing gets deprecated, nothing gets removed.

Bishop: Does anyone agree with Everson?

Whistler: Mansour doesn't want them disunified He's a Copt but not an expert on Coptic. Better argument to disunify than with Cyrrilic and Old Church Slavonic. Practical impact... since Everson strongly believes in disunification, he will come to WG2 to disunify.

Freytag: In WG2 we can say we haven't processed when Everson does present to WG2.

Bishop: Then no motion is necessary.

Ksar: What does disunification here mean?

Whistler: In Unicode Standard we present how to render Coptic. We would need to spell out.

Freytag: A two-thirds vote is needed to disunify.

3.A.4. SPECIFIC SCRIPTS - WHOLE SCRIPTS – NEW PROPOSALS
3.A.4.b. Directionality of Ancient Scripts

Jenkins: Etruscan is left to right by default but can make it right to left by directional overrides. Most scholars are used to seeing left to right.

Sargent: Are there any ligatures?

McGowan: Not to speak of

Freytag: Old scripts general guidelines are: 1) capturing content; and, 2) text based facsimile with varying degrees of faithfulness. Unicode's goal is to encode the legible content of plain text and nothing more. We should look at default layout behavior in the context of modern scholarly writing. Be extremely careful about going into details on our overrides.

McGowan: I agree with the sentiment to address current scholarly usage. I still prefer to stick with the dominant directionality of a script. Need to strike a balance.

Freytag: Does not satisfy re-creation of ancient scripts.

Whistler: Generally, want to agree with Asmus. Where I disagree with Asmus is with the question of where glyph mirroring is expected when you use right-to-left. Script descriptions should mention if glyph mirroring is expected when script is written right-to-left.

Jenkins: It behooves us to pay attention to scholars. Current practice is artifact of technological limitations. Not necessarily a preference.

4. BIDI ISSUES
4.A. PROPOSED CHANGES TO BI-DIRECTIONAL CATEGORIES

4.A.1. Proposal #1 (McConnell) [97-248] FULL STOP, FIGURE SPACE, AMPERSAND, COMMERCIAL AT

Changes to Bi-directional Categories for Full Stop,
Figure Space, Ampersand, Commercial At
[#74-M14] Motion: That the UTC accepts the changes in
bidirectional character types proposed in Table 1 of
document L2/97-248.
Moved by Bishop, seconded by Moore.
Unanimously Approved

4.A.2. Proposal #2 (McConnell) [97-249] SOLIDUS

Whistler: 252 B
Moore: We have received feedback which is reverse of proposal. Most of our users - would like to have dates if we have to choose.

Ksar: As an Arabic expert, my preference is to have date. That is what is useful.

Davis: On "most dates are generated", I think that most are typed in. Some systems don't generate for short dates. The question is frequency of usage. Most agree - dates in text dominate over fractions.

McGowan: It is absurd that fractions would get mixed up.

Davis: Once Unicode is in prominent usage, can use fraction slash. The whole BiDi algorithm is designed to do as good a job as possible. In user data, dates swamp fractions.

Freytag: Part of the problem could be removed if we had a better BiDi algorithm. Dates you can overcome, but can't do so for fractions. Problem is very complex.

Bishop: Fraction slash vs. Solidus. Like it or not there is a lot of 8-bit code out there

Davis: Fraction slash should give you the fraction.

McGowan: Real world has an ambiguous slash.

Davis: The fundamental question - In how many cases are slashes used for dates and how many used for fractions?

Uma: Change the algorithm itself rather than the classification.

Texin: I would separate the algorithm. Build in assumptions about legacy data. We can't decide by frequencies or personal opinion. Pick one and do the work that falls out.

Freytag: Users have the right to make a demand the way they want.

Whistler: Here is a case where we have existing behavior, has been in algorithm for some time and I'm not hearing consensus. Safest thing is to leave it as it is

Moore: Some feedback from Eastern Character Group - they prefer dates format preserved.

Changes to Bi-directional Category for Solidus
[#74-M15] Motion: That the UTC accepts the change to bi-directional character type proposed for the character SOLIDUS in document L2/97-249R..
Moved by Freytag, seconded by McGowan.
Motion failed: 3 for, 9 opposed, 2 abstentions

Aliprand: Change algorithm itself.

Davis: That is more dangerous.

Bishop: #250R Table in Chapter 3, normative, shown as mirrored. Even if rare occurrence, we need to specify how every character behaves in BiDi.

Whistler: Implies a change in 10646?

Bishop: No, for 10646 is an informative annex.

Corrigendum to Algorithm re Depth of Stack [97-257]
[#74-M16] Motion: That the UTC authorizes the addition of corrigenda to the BiDi algorithm based upon text in document L2/97-257.
Moved by Davis, seconded by Bishop.

Freytag: Normatively, after 15 levels.

Davis: To get the same result on all machines.

Bishop: Clarification of current text. Want clearly defined. Beyond 15 levels, the BiDi algorithm doesn't do anything with it.

Davis: Only question is: does it screw up your strings?

Roberts: How large a number do I have to support to be Unicode compliant?

Davis: 32 bit integer. .... If you do go past 16, ignore match.

Amended by Freytag: Retain the statement that, normatively, behavior is undefined. Add a suggestion "how to handle" for this condition.
Corrigendum to Algorithm re Depth of Stack [97-257]
[#74-M16] Motion: That the UTC authorizes the addition of corrigenda to the BiDi algorithm based upon text in document L2/97-257.
Moved by Davis, seconded by Bishop.
Approved as amended: 10, 1, 3

[Oracle left, 13 members present]

Action 74-30 for Davis: Prepare corrigenda to BIDI algorithm based on L2/97-257

Freytag: 3.15 "only levels 0-15 are valid." We are talking about the behavior of the algorithm.

Whistler: There is a hole in the algorithm, we need to define what to do.

Freytag: We should say "only explicit levels 0-15 are valid" Distinguish between levels and push and pop characters.

Whistler: Trying to plug hole in algorithm. "If the level is higher than 16, don't set the value in the array."

4.B.BIDI issues - CORRIGENDA
4.B.1.

[#74-M17] Motion: That the UTC accepts the changes to
bi-directional properties for the control characters
U+000A through U+001F, as proposed in document L2/97-259.
Moved by Bishop, seconded by Whistler.
Approved 12, 0, 1

Bishop: I don't feel strongly about other changes, but good form.

Davis: Very leery about making lots of changes since some have already been implemented.

Freytag: We need to have list broken down, to parse into parts.

Whistler: Opportunity for 2.1 publications.

Davis: If we change, we make a lot of people's code invalid.

Whistler: Clarify. This came as an exercise. Became apparent that BiDi properties were not carefully reviewed.

Freytag: Technical corrigenda... this is appropriate.

Davis: Is this enough of an improvement to change the algorithm? What really has to be changed, come up with a list.

Bishop: Provide reference algorithm. Signed license when it comes time to make it normative. How do we test algorithm x against the reference.

Freytag: Up to your testers to run test cases. We don't certify.

Ksar: Point that Bishop is bringing up - Why does reference implementation have to be normative not informative?

Freytag: Not correct one unless normative.

Bishop: What matters is the behavior. Reference algorithm is a reference for behavior.

Ksar: Avenue is put it into a Technical Report.

Bishop: I want - in short term - for it to be informative.

BiDi Reference Algorithm
[#74-M18] Motion: That the UTC authorizes production
of Unicode Technical Report #9, to consist of a short

description of the BiDi algorithm, with the bulk
of the document to be the algorithm itself.
Moved by Freytag, seconded by Ksar.
Approved 12, 0, 0

Action 74-31 for Bishop: Prepare introductory text of UTR #9 (BiDi Reference Algorithm) and
template by January 31, 1998.
Action 74-32 for Freytag: Assist Bishop in preparation of UTR#9 (BiDi Reference Algorithm)
Action 74-33 for Bishop and Freytag: Send UTR #9 draft to Winkler for distribution by January
31, 1998.
Action 74-34 for Aliprand/Winkler: Put UTR #9 on agenda for February UTC/L2 joint meeting.

Bishop: People - please take a look at this, because people will be using this for at least five years.

[Return to Agenda item]
3.C.3. Cartouche proposal for keyboard symbols (Freytag)

Freytag: L2/97-206 Cartouche - Reference images are not on list, but Michel Suignard will do.
Resolution is to create proposal.

Cartouche Proposal for Keyboard Symbols [97-206]
First Motion:
To accept document L2/97-206.
Moved by Umamaheswaran, seconded by Whistler.
Motion failed: 3 for, 3 against, 7 abstentions

Uma: Split into two

Freytag: Not a friendly amendment. Maybe Unicode might need a cartouche character set,
especially for hieroglyphics.

Whistler: Why do we want to encode? Basically a defensive move to protect against a whole
bunch of discrete characters.

Ksar: I am in favor of accepting but at the same time don't want to spend more time. This is minor,
let's not waste too much more time.

Freytag: Don't need to resolve all cartouching issues now. Can take this proposal.

Second motion:
To accept the character COMBINING ENCLOSING SCREEN
CARTOUCHE.
Moved by Umamaheswaran, seconded by Suignard
Motion failed: 6 for, 1 against, 6 abstentions

Suignard: We have an immediate use.

Roberts: We wanted to see what the proposal would be like.

Uma: Screen cartouche is useful as it is. Carry forward to WG2.

[#74-M19] **Motion:** That the UTC accepts proposal
97-206, with editorial clarification.
Moved by McGowan, seconded by Freytag.
Approved 7, 2, 4

5. CURRENT BALLOTS
6. W3C ISSUES

UTC Representation at IRG meeting 10
[#74-M20] Motion: That the UTC appoints Hideki
Hiura to attend IRG #10 in Ho-Chi Minh City,
Vietnam as an observer from the Unicode Consortium,
but without voting authority. In nay vote, he
should vote "not present."
Moved by Freytag, seconded by Jenkins.
Unanimously Approved

Action 74-42 for Jenkins: Talk with Mr. Zhong regarding Hidecki's status as an observer at the
IRG in Ho-Chi Min City
Action 74-41 for Aliprand: Talk with UTC observer, Mr. Hidecki.

UTC Representation at Cultural Adaptability Workshop
[#74-M21] Motion: That the UTC appoints Asmus
Freytag to represent the Unicode Consortium at
the Cultural Adaptability Workshop in Ottawa,
Canada, January 20-22, 1998.
Moved by Winkler, seconded by O'Donnell.
Approved by Consensus

7. STANDING ITEMS
　　7.A. International Unicode Conference

8.　OTHER BUSINESS

9.　INTERACTION WITH SC2, PART 2

10. See Document L2/97-282R

11.REVIEW OF MOTIONS TO L2, AND ACTION ITEMS

12.CLOSING OF JOINT MEETING

Name	Company	L2 12/3/97	UTC 12/3/97	E-mail
John Jenkins	Apple	P	CM	jenkins@apple.com
Rick McGowan	Apple	x	CM	rmcgowan@apple.com
Sandra Martin O'Donnell	Digital	P	CM	odonnell@zk3.dec.com
Don Carroll	Hewlett-Packard	A	CM	mike_ksar@hp.com
Mike Ksar	Hewlett-Packard	P	CM	dcarroll@sea.hp.com
Mark Davis	Taligent/IBM		P, CM	Mark@unicode.org
Lisa Moore	IBM	A	VP, CM	lisam@us.ibm.com
V.S. Umamaheswaran	IBM	P	CM	umavs@ca.ibm.com
Tatsuo Kobayashi	Justsystem		CM	tatsuo_kobayashi@justsystem.co.jp
F. Avery Bishop	Microsoft	x	CM	averyb@microsoft.com
Mike Kernaghan	Microsoft	x	VP, CM	mikekern@microsoft.com
Murray Sargent	Microsoft	P	CM	murrays@microsoft.com
Michel Suignard	Microsoft	A	CM	michelsu@microsoft.com
Gary Roberts	NCR	X	CM	gary.roberts@sandiegoca.ncr.com
Lloyd Honomichl	Novell	P	CM	lloyd_honomichl@novell.com
Chris Boyle	Novell	O		dcboyle@novell.com
Tex Texin	Progress	P	AM	texin@progress.com
Nelson Ng	Oracle	P		nng@us.oracle.com
Joan Aliprand	RLG	P	CM	br.jma@rlg.org
Hideki Hiura	Sun Microsystems	P	CM	hiura@eng.sun.com
Ken Whistler	Sybase	P	CM	kenw@sybase.com
Asmus Freytag	Unicode	P	VP, AM	asmusf@ix.netcom.com
Julie Allen	Unicode	O	O	julie@unicode.org
Julia Oesterle	Unicode	O	O	julia@unicode.org
Arnold Winkler	Unisys	P	CM	arnold.winkler@unisys.com
Joe Becker	Xerox	P	CM	becker.osbu_north@xerox.com
John Fiscella	Production First Software	O	x	profirst@compuserve.com
Allain La Bonté	Governement du Québec	O	x	alb@riqu.qc.ca

Legend: P - primary, A - alternate, O - observer, L - liaison, X - ex-officio, x - present
CM - corporate member, AM - associate member, VP - vice president, Unicode, Inc.